

09-05-00

# HAMILTON, BROOK, SMITH & REYNOLDS, P.C.

## UTILITY PATENT APPLICATION TRANSMITTAL

(Only for new nonprovisional applications under  
37 C.F.R. 1.53(b))

Attorney Docket No.

0918.1305-000

First Named Inventor or  
Application Identifier

Vladimir Pavlović

Express Mail Label No.

EL 551544800 US

Title of  
Invention

METHOD FOR MOTION SYNTHESIS AND INTERPOLATION USING SWITCHING  
LINEAR DYNAMIC SYSTEM MODELS

### APPLICATION ELEMENTS

See MPEP chapter 600 concerning utility patent application contents.

ADDRESS TO:

Assistant Commissioner for Patents  
Box Patent Application  
Washington, D.C. 20231

1. ☒ Fee Transmittal Form  
(Submit an original, and a duplicate for fee processing)
2. ☒ Specification **[Total Pages [76]]**  
(preferred arrangement set forth below)
  - Descriptive title of the invention
  - Cross References to Related Applications
  - Statement Regarding Fed sponsored R & D
  - Reference to microfiche Appendix
  - Background of the Invention
  - Summary of the Invention
  - Brief Description of the Drawings
  - Detailed Description
  - Claim(s)
  - Abstract of the Disclosure
3. ☒ Drawing(s) (35 U.S.C. 113) **[Total Sheets [21]]**  
[ ] Formal [X] Informal
4. ☒ Oath or Declaration **[Total Pages [1]]**
  - a. ☒ Newly executed (original or copy)
  - b. [ ] Copy from a prior application (37 C.F.R. 1.63(d))  
(for continuation/divisional with Box 17 completed)  
**[NOTE Box 5 below]**
    - i. [ DELETION OF INVENTOR(S) ] Signed statement attached deleting inventor(s) named in the prior application, see 37 C.F.R. 1.63(d)(2) and 1.33(b).
5. [ ] Incorporation By Reference (useable if Box 4b is checked)  
The entire disclosure of the prior application, from which a copy of the oath or declaration is supplied under Box 4b, is considered as being part of the disclosure of the accompanying application and is hereby incorporated by reference therein.

6. [ ] Microfiche Computer Program (Appendix)
7. [ ] Nucleotide and/or Amino Acid Sequence Submission (if applicable, all necessary)
  - a. [ ] Computer Readable Copy
  - b. [ ] Paper Copy (identical to computer copy)  
[ ] Pages
  - c. [ ] Statement verifying identity of above copies

### ACCOMPANYING APPLICATION PARTS

8. ☒ Assignment Papers (cover sheet & documents)
9. [ ] 37 C.F.R. 3.73(b) Statement [ ] Power of Attorney  
(when there is an assignee)
10. [ ] English Translation Document (if applicable)
11. [ ] Information Disclosure Statement (IDS)/PTO-1449 [ ] Copies of IDS Citations
12. [ ] Preliminary Amendment
13. ☒ Return Receipt Postcard (MPEP 503)  
(Should be specifically itemized)
14. [ ] Small Entity [ ] Statement filed in prior application, status still proper and desired
15. [ ] Certified Copy of Priority Document(s)  
(if foreign priority is claimed)
16. ☒ Other: Power of Attorney by Assignee

17. If a CONTINUING APPLICATION, check appropriate box and supply the requisite information:

[ ] Continuation [ ] Divisional [ ] Continuation-in-part (CIP) of prior application No.:

Prior application information: Examiner:

Group Art Unit:

### 18. CORRESPONDENCE ADDRESS

NAME	James M. Smith, Esq.				
	HAMILTON, BROOK, SMITH & REYNOLDS, P.C.				
ADDRESS	Two Militia Drive				
CITY	Lexington	STATE	MA	ZIP CODE	02421-4799
COUNTRY	USA	TELEPHONE	(781) 861-6240	FAX	(781) 861-9540
Signature	<i>Gerald M. Bluhm</i>			Date	9/1/00
Submitted by Typed or Printed Name	Gerald M. Bluhm			Reg. Number	44,035

# HAMILTON, BROOK, SMITH & REYNOLDS, P.C.

## FEE TRANSMITTAL FOR PATENT APPLICATIONS

Attorney Docket Number	0918.1305-000
Application Number	
First Named Inventor	Vladimir Pavlović

JEP 918 U.S. PTO  
 09/654401  
 09/01/00

CLAIM CALCULATION (includes any preliminary amendment)

CLAIMS	(1) FOR	(2) NUMBER FILED	(3) NUMBER EXTRA	(4) RATE	(5) CALCULATIONS
	TOTAL CLAIMS (37 CFR 1.16(c) or (j))	74 - 20* =	54	x \$ 18 =	\$ 972
	INDEPENDENT CLAIMS (37 CFR 1.16(b) or (i))	16 - 3** =	13	x \$ 78 =	\$ 1014
	MULTIPLE DEPENDENT CLAIMS (if applicable) (37 CFR 1.16(d))			+ \$ 260 =	\$
				BASIC FEE (37 CFR 1.16(a) or (h))	\$ 690
	Total of above Calculations =				\$ 2676
	Reduction by 50% for filing by small entity (37 CFR 1.9, 1.27, 1.28) =				\$
	TOTAL =				\$ 2676
	Surcharge - Late Filing of Declaration or Filing Fees (37 C.F.R. 1.16(e)) =				\$
	Petition for Extension of Time Fee (37 C.F.R. 1.17) =				\$
	Assignment Recordation Fee = (only when filed with application)				\$ 40
	TOTAL =				\$ 2716

\* Reissue claims in excess of 20 and over original patent  
 \*\* Reissue independent claims over original patent

1. Small entity status:

- a. ☐ A small entity statement is enclosed.
- b. ☐ A small entity statement was filed in the prior non-provisional application and such status is still proper and desired.
- c. ☐ Is no longer claimed.

2. ☒ A general authorization is hereby granted to charge deposit account number 08-0380 for any fees required under 37 CFR 1.16 and 1.17 in order to maintain pendency of this application. A copy of this authorization is enclosed for accounting purposes.

3. ☒ A check is enclosed for \$2716. ☐ Please charge \$[ ] to Deposit Account No. 08-0380.

4. ☐ Other: \_\_\_\_\_

Signature	<i>Gerald M. Bluhm</i>	Date	9/1/00
Submitted by Typed or Printed Name	Gerald M. Bluhm	Reg. Number	44,035

-1-

Date: 9-1-00 Express Mail Label No. EL 5515 44800US

Inventor: Vladimir Pavlović and James M. Rehg  
Attorney's Docket No.: 0918.1305-000

# METHOD FOR MOTION SYNTHESIS AND INTERPOLATION USING SWITCHING LINEAR DYNAMIC SYSTEM MODELS

## RELATED APPLICATIONS

This Application claims the benefit of U.S. Provisional Application No.

5 60/154,384, filed September 16, 1999, the entire teachings of which are incorporated  
herein by reference.

## BACKGROUND OF THE INVENTION

Technologies for analyzing the motion of the human figure play a key role in a broad range of applications, including computer graphics, user-interfaces, surveillance, and video editing.

A motion of the figure can be represented as a trajectory in a state space which is defined by the kinematic degrees of freedom of the figure. Each point in state space represents a single configuration or pose of the figure. A motion such as a pli   in ballet is described by a trajectory along which the joint angles of the legs and arms change continuously.

A key issue in human motion analysis and synthesis is modeling the dynamics of the figure. While the kinematics of the figure define the state space, the dynamics define which state trajectories are possible (or probable).

Since the key problem in synthesizing figure motion for animation is to achieve realistic dynamics, the importance of dynamic modeling is obvious. The challenge in animation is to produce motion with natural dynamics that satisfy constraints placed by

the animator. Some constraints result from basic physical realities such as the noninterpenetration of objects. Others are artistic in nature, such as a desired head pose during a dance move. The key problem in synthesis is to find a trajectory in the set of dynamically realistic trajectories that satisfies the desired constraints.

## 5 Prior Approaches

Most previous work on synthesizing figure motion employs one of two types of dynamic models: analytic and learned. Analytic models are specified by a human designer. They are typically second order differential equations relating joint torque, mass, and acceleration. Learned models, on the other hand, are constructed automatically from examples of human motion data.

### Analytic Dynamic Models

The prior art includes a range of hand-specified analytic dynamical models. On one end of the spectrum are simple generic dynamic models based, for example, on constant velocity assumptions. Complex, highly specific models occupy the other end.

15 A number of proposed figure trackers use a generic dynamic model based on a simple smoothness prior such as a constant velocity Kalman filter. See, for example, Ioannis A. Kakadiaris and Dimitris Metaxas, "Model-based estimation of 3D human motion with occlusion based on active multi-viewpoint selection," Computer Vision and pattern Recognition, pages 81-87, San Francisco, CA, June 18-20, 1996. Such

20 models fail to capture subtle differences in dynamics across different motion types, such as walking or running. It is unlikely that these models can provide a strong constraint on complex human motion such as dance.

The field of biomechanics is a source of more complex and realistic models of human dynamics. From the biomechanics point of view, the dynamics of the figure are

25 the result of its mass distribution, joint torques produced by the motor control system, and reaction forces resulting from contact with the environment, e.g., the floor. Research efforts in biomechanics, rehabilitation, and sports medicine have resulted in

complex, specialized models of human motion. For example, entire books have been written on the subject of walking. See, for example, Inman, Ralston and Todd, "Human Walking," Williams and Wilkins, 1981.

The biomechanical approach has two drawbacks for analysis and synthesis applications. First, the dynamics of the figure are quite complex, involving a large number of masses and applied torques, along with reaction forces which are difficult to measure. In principle, all of these factors must be modeled or estimated in order to produce physically-valid dynamics. Second, in some applications we may only be interested in a small set of motions, such as a vocabulary of gestures. In the biomechanical approach, it may be difficult to reduce the complexity of the model to exploit this restricted focus. Nonetheless, these models have been applied to tracking and synthesis applications.

Wren and Pentland, "Dynamic models of human motion", Proceeding of the Third International Conference on Automatic Face and Gesture Recognition, pages 22-27, Nara, Japan, 1998, explored visual tracking using a biomechanically-derived dynamic model of the upper body. The unknown joint torques were estimated along with the state of the arms and head in an input estimation framework. A Hidden Markov Model (HMM) was trained to represent plausible sequences of input torques. Due to the simplicity of their experimental domain, there was no need to model reaction forces between the figure and its environment.

This solution suffers from the limitations of the biomechanical approach outlined above. In particular, describing the entire body would require a significant increase in the complexity of the model. Even more problematic is the treatment of the reaction forces, such as those exerted by the floor on the soles of the feet during walking or running.

Biomechanically-derived dynamic models have also been applied to the problem of synthesizing athletic motion, such as bike racing or sprinting, for computer graphics animations. See, for example, Hodgins, Wooten, Brogan and O'Brien, "Animating human athletics," Computer Graphics (Proc. SIGGRAPH '95), pages 71-

78, 1995. In the present invention, there is, in addition to the usual problems of complex dynamic modeling, the need to design control programs that produce the joint torques that drive the figure model. In this approach, it is difficult to capture more subtle aspects of human motion without some form of automated assistance. The motions that result tend to appear very regular and robotic, lacking both the randomness and fluidity associated with natural human motion.

#### Learned Dynamic Models

The approaches to figure motion synthesis using learned dynamic models are based on synthesizing motion using dynamic models whose parameters are learned from a corpus of sample motions.

In Brand, "Pattern Discovery via Entropy Minimization," Technical Report TR98-21, Mitsubishi Electric Research Lab, 1998, an HMM-based framework for dynamics learning is proposed and applied to synthesis of realistic facial animations from a training corpus. The main component of this work is the use of an entropic prior to cope with sparse input data.

Brand's approach has two potential disadvantages. First, it assumes that the resulting dynamic model is time invariant; each state space neighborhood has a unique distribution over state transitions. Second, the use of entropic priors results in fairly "deterministic" models learned from a moderate corpus of training data. In contrast, the diversity of human motion applications require complex models learned from a large corpus of data. In this situation, it is unlikely that a time invariant model will suffice, since different state space trajectories can originate from the same starting point depending upon the class of motion being performed.

#### Motion Capture for Motion Synthesis

A final category of prior art which is relevant to this invention is the use of motion capture to synthesize human motion with realistic dynamics. Motion capture is by far the most successful commercial technique for creating computer graphics

animations of people. In this method, the motion of human actors is captured in digital form using a special suit with either optical or magnetic sensors or targets. This captured motion is edited and used to animate graphical characters.

5 The motion capture approach has two important limitations. First, the need to wear special clothing in order to track the figure limits the application of this technology to motion which can be staged in a studio setting. This rules out the live, real-time capture of events such as the Olympics, dance performances, or sporting events in which some of the finest examples of human motion actually occur.

10 The second limitation of current motion capture techniques is that they result in a single prototype of human motion which can only be manipulated in a limited way without destroying its realism. Using this approach, for example, it is not possible to synthesize multiple examples of the same type of motion which differ in a random fashion. The result of motion capture in practice is typically a kind of "wooden", fairly inexpressive motion that is most suited for animating background characters. That is  
15 precisely how this technology is currently used in Hollywood movie productions.

There is a clear need for more powerful tracking techniques that can recover human motion under less restrictive conditions. Similarly, there is a need for more powerful generative models of human motion that are both realistic and capable of generating sample motions with natural amounts of "randomness."

## 20 SUMMARY OF THE INVENTION

Technologies for analyzing the motion of the human figure play a key role in a broad range of applications, including computer graphics, user-interfaces, surveillance, and video editing. A motion of the figure can be represented as a trajectory in a state space which is defined by the kinematic degrees of freedom of the figure. Each point in  
25 state space represents a single configuration or pose of the figure. A motion such as a pli   in ballet is described by a trajectory along which the joint angles of the legs and arms change continuously.

007050-1045950

5 models. Analytic models are specified by a human designer. They are typically second order differential equations relating joint torque, mass, and acceleration.

10 system, and reaction forces resulting from contact with the environment (e.g. the floor).  
Research efforts in biomechanics, rehabilitation, and sports medicine have resulted in  
complex, specialized models of human motion. For example, detailed walking models  
are described in Inman et al., “Human Walking,” Williams and Wilkins, 1981.

figure are quite complex, involving a large number of masses and applied torques, along with reaction forces which are difficult to measure. In principle all of these factors must be modeled or estimated in order to produce physically-valid dynamics. Second, in some applications we may only be interested in a small set of motions, such as a vocabulary of gestures. In the biomechanical approach it may be difficult to reduce the complexity of the model to exploit this restricted focus. Nonetheless, biomechanical models have been applied to human motion analysis.

25 Recognition, pages 22-27, Nara, Japan, 1998. The unknown joint torques are estimated along with the state of the arms and head in an input estimation framework. A Hidden Markov Model is trained to represent plausible sequences of input torques. This prior art does not address the problem of modeling reaction forces between the figure and its



environment. An example is the reaction force exerted by the floor on the soles of the feet during walking or running.

Therefore, there is a need for inference and learning methods for fully coupled SLDS models that can estimate a complete set of model parameters for a switching  
5 model given a training set of time-series data.

Described herein is a new class of approximate learning methods for switching linear dynamic (SLDS) models. These models consist of a set of linear dynamic system (LDS) models and a switching variable that indexes the active model. This new class has three advantages over dynamics learning methods known in the prior art:

- 10       \*     New approximate inference techniques lead to tractable learning even when the set of LDS models is fully coupled.
- \*     The resulting models can represent time-varying dynamics, making them suitable for a wide range of applications.
- \*     All of the model parameters are learned from data, including the plant  
15       and noise parameters for the LDS models and Markov model parameters for the switching variable.

In addition, this method can be applied to the problem of learning dynamic models for human motion from data. It has three advantages over analytic dynamic  
20 models known in the prior art:

- \*     Models can be constructed without a laborious manual process of specifying mass and force distributions. Moreover, it may be easier to tailor a model to a specific class of motion, as long as a sufficient number of samples are available.
- 25       \*     The same learning approach can be applied to a wide range of human motions from dancing to facial expressions.
- \*     When training data is obtained from analysis of video measurements, the spatial and temporal resolution of the video camera determine the level of detail at which dynamical effects can be observed. Learning

techniques can only model structure which is present in the training data. Thus, a learning approach is well-suited to building models at the correct level of resolution for video processing and synthesis.

- A wide range of learning algorithms can be cast in the framework of Dynamic Bayesian Networks (DBNs). DBNs generalize two well-known signal modeling tools: Kalman filters for continuous state linear dynamic systems (LDS) and Hidden Markov Models (HMMs) for discrete state sequences. Kalman filters are described in Anderson et al., "Optimal Filtering," Prentice-Hall, Inc., Englewood Cliffs, NJ, 1979. Hidden Markov Models are reviewed in Jelinek, "Statistical methods for speech recognition" MIT Press, Cambridge, MA, 1998.

- Dynamic models learned from sequences of training data can be used to predict future values in a new sequence given the current values. They can be used to synthesize new data sequences that have the same characteristics as the training data. They can be used to classify sequences into different types, depending upon the conditions that produced the data.

- We focus on a subclass of DBN models called Switching Linear Dynamics Systems. Intuitively, these models attempt to describe a complex nonlinear dynamic system with a succession of linear models that are indexed by a switching variable. The switching approach has an appealing simplicity and is naturally suited to the case where the dynamics are time-varying.

We present a method for approximate inference in fully coupled switching linear dynamic models (SLDSs). Exponentially hard exact inference is replaced with approximate inference of reduced complexity.

- The first preferred embodiment uses Viterbi inference jointly in the switching and linear dynamic system states.

The second preferred embodiment uses variational inference jointly in the switching and linear dynamic system states.

The third preferred embodiment uses general pseudo Bayesian inference jointly in the switching and linear dynamic system states.

Parameters of a fully connected SLDS model are learned from data. Model parameters are estimated using a generalized expectation-maximization (EM) algorithm. Exact expectation/inference (E) step is replaced with one of the three approximate inference embodiments.

The learning method can be used to model the dynamics of human motion. The joint angles of the limbs and pose of the torso are represented as state variables in a switching linear dynamic model. The switching variable identifies a distinct motion regime within a particular type of human motion.

Motion regimes learned from figure motion data correspond to classes of human activity such as running, walking, etc. Inference produces a single sequence of switching modes which best describes a motion trajectory in the figure state space. This sequence segments the figure motion trajectory into motion regimes learned from data.

Accordingly, a method for synthesizing a sequence includes defining a switching linear dynamic system (SLDS) having a plurality of dynamic models, where each model is associated with a switching state such that a model is selected when its associated switching state is true. A state transition record for one or more training  
20 sequence of measurements is determined by determining and recording, for a given measurement and for each possible switching state, an optimal prior switching state, based on the training sequences, where the optimal prior switching state optimizes a transition probability. An optimal final switching state is then determined for a final measurement. Next, the sequence of switching states is determined by backtracking  
25 through the state transition record, starting from the optimal final switching state. Parameters of the dynamic models are learned in response to the determined sequence of switching states. Finally, a new data sequence is synthesized, based on the dynamic models whose parameters have been learned.

The new data sequence can have characteristics which are similar to characteristics of at least one training sequence. Alternatively, the new data sequence can combine characteristics of plural training sequences which have different characteristics.

- 5 In at least one embodiment, the SLDS is modified such that one or more constraints are met. This modification can be accomplished, for example, by adding a continuous state control, such as one or more constraints on continuous states, constraints on the continuous state control, and/or constraints on time. An optimal continuous control can be designed that satisfies the at least one constraint. In the latter
- 10 case, the new data sequence can be synthesized using the optimal control.

Alternatively, the SLDS modification can be accomplished by adding a switching state control, such as one or more constraints on switching states and/or constraints on the switching state control.

- Both optimal switching and continuous state controls can be designed that
- 15 satisfy continuous and switching constraints respectively.

In various embodiments of the present invention, the sequence of measurements can comprise, but is not limited to, economic data, image data, audio data and/or spatial data.

- In one embodiment of the present invention, a SLDS model includes a state
- 20 transition recorder which determines the state transition record, and which determines the optimal final switching state for the a final measurement. A backtracker determines the sequence of switching states corresponding to the training sequence by backtracking, from the optimal final switching state, through the state transition record. A dynamic model learner learns parameters of the dynamic models responsive to the
- 25 determined sequence of switching states, and a synthesizer synthesizes a new data sequence, based on dynamic models whose parameters have been learned.

While the above embodiments are based on Viterbi techniques, other embodiments of the present invention are based on variational techniques. For example, method for synthesizing a sequence includes defining a switching linear

0054401-090100

dynamic system (SLDS) having a plurality of dynamic models. Each dynamic model is associated with a switching state such that a dynamic model is selected when its associated switching state is true. The switching state at a particular instance is determined by a switching model, such as a hidden Markov model (HMM). The

5 dynamic models are decoupled from the switching model, and parameters of the decoupled dynamic model are determined responsive to a switching state probability estimate. A state of a decoupled dynamic model corresponding to a measurement at the particular instance is estimated, responsive to one or more training sequences. Parameters of the decoupled switching model are then determined, responsive to the

10 dynamic state estimate. A probability is estimated for each possible switching state of the decoupled switching model. The sequence of switching states is determined based on the estimated switching state probabilities. Parameters of the dynamic models are learned responsive to the determined sequence of switching states. Finally, a new data sequence is synthesized based on the dynamic models with learned parameters.

15 A switching linear dynamic system (SLDS) model based on variational techniques includes an approximate variational state sequence inference module, which reestimates parameters of each SLDS model, using variational inference, to minimize a modeling cost of current state sequence estimates, responsive to one or more training measurement sequences. A dynamic model learner learns parameters of the dynamic

20 models responsive to the determined sequence of switching states, and a synthesizer synthesizes a new data sequence, based on the dynamic models with learned parameters.

Another embodiment of the present invention comprises a method for interpolating from a measurement sequence, and includes defining a SLDS having a

25 plurality of dynamic models. Each model is associated with a switching state such that a model is selected when its associated switching state is true. A state transition record is determined by determining and recording, for a given measurement and for each possible switching state, an optimal prior switching state, based on the measurement sequence, where the optimal prior switching state optimizes a transition probability. An



Fig. 6 is a dependency graph illustrating the decoupling of the hidden Markov model and SLDS in the embodiment of Fig. 5.

Fig. 7 is a flowchart illustrating the steps performed by the embodiment of Fig. 5.

5 Figs. 8A and 8B comprise a flowchart illustrating the steps performed by a GPB2 embodiment of the present invention.

Fig. 9 comprises two graphs which illustrate learned segmentation of a "jog" motion sequence.

Fig. 10 is a block diagram illustrating classification of state space trajectories, as  
10 performed by the present invention.

Fig. 11 comprises several graphs which illustrate an example of segmentation.

Fig. 12 is a block diagram of a Kalman filter as employed by an embodiment of the present invention.

Fig. 13 is a diagram illustrating the operation of the embodiment of Fig. 12 for  
15 the specific case of figure tracking.

Fig. 14 is a diagram illustrating the mapping of templates.

Fig. 15 is a block diagram of an iterated extended Kalman filter (IEKF).

Fig. 16 is a block diagram of an embodiment of the present invention using an IEKF in which a subset of Viterbi predictions is selected and then updated.

20 Fig. 17 is a block diagram of an embodiment in which Viterbi predictions are first updated, after which a subset is selected.

Fig. 18 is a block diagram of an embodiment which combines SLDS prediction with sampling from a prior mixture density.

Fig. 19 is a block diagram of an embodiment in which Viterbi estimates are  
25 combined with updated samples drawn from a prior density.

Fig. 20 is a block diagram illustrating an embodiment of the present invention in which the framework for synthesis of state space trajectories in which an SLDS model is used as a generative model.

Fig. 21 is a dependency graph of an SLDS model, modified according to an embodiment of the present invention, with added continuous state constraints.

Fig. 22 is a dependency graph of an SLDS model, modified according to an embodiment of the present invention, with added switching state constraints.

5 Fig. 23 is a dependency graph of an SLDS model, modified according to an embodiment of the present invention, with both added continuous and switching state constraints.

Fig. 24 is a block diagram of the framework for a synthesis embodiment of the present invention using constraints and utilizing optimal control.

10 Fig. 25 is an illustration of a stick figure motion sequence as synthesized by an embodiment of the present invention.

## DETAILED DESCRIPTION OF THE INVENTION

### SWITCHING LINEAR DYNAMIC SYSTEM MODEL

Fig. 1 is a block diagram of a complex physical linear dynamic system (LDS) 10, driven by white noise  $v_k$ , also called “plant noise.” The LDS state parameters evolve in time according to some known model, such as a Markov chain (MC) model 12.

The system can be described using the following set of state-space equations:

$$\begin{aligned} x_{t+1} &= A(s_{t+1})x_t + v_{t+1}(s_{t+1}), \\ y_t &= C x_t + w_t, \text{ and} \\ x_0 &= v_0(s_0) \end{aligned} \quad (\text{Eq. 1})$$

20 for the physical system 10, and

$$\begin{aligned} \Pr(s_{t+1}|s_t) &= s'_{t+1} \prod s_t, \text{ and} \\ \Pr(s_0) &= \pi_0 \end{aligned}$$

for the switching model 12.



Here,  $x_t \in \mathcal{R}^N$  denotes the hidden state of the LDS 10 at time  $t$ , and  $v_t$  is the state noise process. Similarly,  $y_t \in \mathcal{R}^M$  is the observed measurement at time  $t$ , and  $w_t$  is the measurement noise. Parameters  $A$  and  $C$  are the typical LDS parameters: the state transition matrix 14 and the observation matrix 16, respectively. Assuming the

5 LDS models a Gauss-Markov process, the noise processes are independently distributed Gaussian:

$$\begin{aligned} v_t(s_t) &\sim N(0, Q(s_t)), t > 0 \\ v_0(s_0) &\sim N(x_0(s_0), Q_0(s_0)) \\ w_t &\sim N(0, R). \end{aligned}$$

where  $Q$  is the state noise variance and  $R$  is the measurement noise variance. The notation  $s'$  is used to indicate the transpose of vector  $s$ .

10 The switching model 12 is assumed to be a discrete first-order Markov process. State variables of this model are written as  $s_t$ . They belong to the set of  $S$  discrete symbols  $\{e_0, \dots, e_{S-1}\}$ , where  $e_i$  is, for example, the unit vector of dimension  $S$  with a non-zero element in the  $i$ -th position. The switching model 12 is a first-order discrete Markov model defined with the state transition matrix  $\Pi$  whose elements are

$$15 \quad \Pi(i, j) = \Pr(s_{t+1} = e_i | s_t = e_j), \quad (\text{Eq. 2})$$

and which is given an initial state distribution  $\Pi_0$ .

Coupling between the LDS and the switching process is full and stems from the dependency of the LDS parameters  $A$  and  $Q$  on the switching process state  $s_t$ . Namely,

$$\begin{aligned} A(s_t = e_i) &= A_i \\ Q(s_t = e_i) &= Q_i \end{aligned}$$

20 In other words, switching state  $s_t$  determines which of  $S$  possible models  $\{(A_0, Q_0), \dots, (A_{S-1}, Q_{S-1})\}$  is used at time  $t$ .

Fig. 2 is a dependency graph 20 equivalently illustrating a rather simple but fully coupled Bayesian network representation of the SLDS, where each  $s_t$  denotes an instance of one of the discrete valued action states which switch the physical system models having continuous valued states  $x$  and producing observations  $y$ . The full

5 model can be written as the “joint distribution”  $P$ :

$$P(Y_T, X_T, S_T) = \Pr(s_0) \prod_{t=1}^{T-1} \Pr(s_t | s_{t-1})$$

$$\Pr(x_0 | s_0) \prod_{t=1}^{T-1} \Pr(x_t | x_{t-1}, s_t)$$

$$\prod_{t=0}^{T-1} \Pr(y_t | x_t).$$

where  $Y_T$ ,  $X_T$ , and  $S_T$  denote the sequences, of length  $T$ , of observation and hidden state variables, and switching variables, respectively. For example,  $Y_T = \{y_0, \dots, y_{T-1}\}$ . In this dependency graph 20, the coupling between the switching states and the LDS states

10 is full, i.e., switching states are time-dependent, LDS states are time-dependent, and switching and LDS states are intradependent.

We can now write an equivalent representation of the fully coupled SLDS as the above DBN, assuming that the necessary conditional probability distribution functions (pdfs) are appropriately defined. Assuming a Gauss-Markov process on the LDS, it

15 follows:

$$x_{t+1} | x_t, s_{t+1} = e_i \sim N(A_i x_t, Q_i),$$

$$y_t | x_t \sim N(Cx_t, R),$$

$$x_0 | s_0 = e_i \sim N(x_{0,i}, Q_0, i)$$

Recalling the Markov switching model assumption, the joint pdf of the complex DBN of duration  $T$ , or, equivalently, its Hamiltonian, where the Hamiltonian  $H(x)$  of a

distribution  $P(x)$  is defined as any positive function such that  $P(x) = [(\exp(-H(x)))/(\sum_{\psi} \exp(-H(\psi)))]$ , can be written as:

$$\begin{aligned}
 H(X_T, S_T, Y_T) = & \frac{1}{2} \sum_{t=1}^{T-1} \sum_{i=0}^{N-1} [(x_t - A_i x_{t-1})' Q_i^{-1} (x_t - A_i x_{t-1}) + \log |Q_i|] s_t(i) \\
 & + \frac{1}{2} \sum_{i=0}^{N-1} [(x_{0,i})' Q_{0,i}^{-1} (x_{0,i}) + \log |Q_{0,i}|] s_0(i) + \frac{NT}{2} \log 2\pi \\
 & + \frac{1}{2} \sum_{t=0}^{T-1} (y_t - Cx_t)' R^{-1} (y_t - Cx_t) + \frac{T}{2} \log |R| + \frac{MT}{2} \log 2\pi \\
 & + \sum_{t=1}^{T-1} s'_t (-\log \prod) s_{t-1} + s'_0 (-\log \pi_0).
 \end{aligned}
 \tag{Eq. 3}$$

## INFERENCE

- 5        The goal of inference in SLDSs is to estimate the posterior probability of the hidden states of the system ( $s_t$  and  $x_t$ ) given some known sequence of observations  $Y_T$  and the known model parameters, i.e., the likelihood of a sequence of models as well as the estimates of states. Namely, we need to find the posterior

$$P(X_T, S_T | Y_T) = \Pr(X_T, S_T | Y_T),$$

- 10    or, equivalently, its “sufficient statistics”. Given the form of  $P$  it is easy to show that these are the first and second-order statistics: mean and covariance among hidden states  $x_t, x_{t-1}, s_t, s_{t-1}$ .

- If there were no switching dynamics, the inference would be straightforward - we could infer  $X_T$  from  $Y_T$  using an LDS inference approach such as smoothing, as described by Rauch, “Solutions to the linear smoothing problem,” IEEE Trans. Automatic Control, AC-8(4):371-372, October 1963. However, the presence of switching dynamics embedded in matrix  $P$  makes exact inference more complicated.

To see that, assume that the initial distribution of  $x_0$  at  $t = 0$  is Gaussian. At  $t = 1$ , the pdf of the physical system state  $x_1$  becomes a mixture of  $S$  Gaussian pdfs since we need to marginalize over  $S$  possible but unknown models. At time  $t$ , we will have a mixture of  $S^t$  Gaussians, which is clearly intractable for even moderate sequence lengths. It is therefore necessary to explore approximate inference techniques that will result in a tractable learning method. What follows are three preferred embodiments of the inference step.

#### APPROXIMATE VITERBI INFERENCE EMBODIMENT

Fig. 3 is a block diagram of an embodiment of a dynamics learning method based on approximate Viterbi inference. At step 32, switching dynamics of a number of SLDS motion models are learned from a corpus of state space motion examples 30. Parameters 36 of each SLDS are re-estimated, at step 34, iteratively so as to minimize the modeling cost of current state sequence estimates, obtained using the approximate Viterbi inference procedure. Approximate Viterbi inference is developed as an alternative to computationally expensive exact state sequence estimation.

The task of the Viterbi approximation approach of the present invention is to find the most likely sequence of switching states  $s_t$  for a given observation sequence  $Y_T$ . If the best sequence of switching states is denoted  $S_T^*$ , then the desired posterior

$P(X_T, S_T | Y_T)$  can be approximated as

$$P(X_T, S_T | Y_T) = P(X_T | S_T, Y_T) P(S_T | Y_T) \approx P(X_T | S_T, Y_T) \delta(S_T - S_T^*), \quad (\text{Eq 4})$$

where  $\delta(x) = 1$  if  $x = 0$  and  $\delta(x) = 0$  if  $x \neq 0$ . In other words, the switching sequence posterior  $P(S_T | Y_T)$  is approximated by its mode. Applying Viterbi inference to two simpler classes of models, discrete state hidden Markov models and continuous state Gauss-Markov models is well-known. An embodiment of the present invention utilizes an algorithm for approximate Viterbi inference that generalizes the two approaches.

We would like to determine the switching sequence  $S_T^*$  such that

$S_T^* = \arg \max_{S_T} P(S_T | Y_T)$ . First, define the following probability  $J_{t,i}$  up to time  $t$  of the switching state sequence being in state  $i$  at time  $t$  given the measurement sequence  $Y_t$ :

$$J_{t,i} = \max_{S_{t-1}} P(S_{t-1}, s_t = e_i, Y_t) \quad (\text{Eq. 5})$$

- 5 If this quantity is known at time  $T$ , the probability of the most likely switching sequence  $S_T^*$  is simply  $P(S_T^* | Y_T) \propto \max_i J_{T,i}$ . In fact, a recursive procedure can be used to obtain the desired quantity. To see that, express  $J_{t,i}$  in terms of  $J_s$  at  $t-1$ . It follows that

$$\begin{aligned} J_{t,i} &= \max_{S_{t-1}} P(S_{t-1}, s_t = e_i, Y_t) \\ &= \max_{S_{t-1}} P(S_{t-1}, s_t = e_i, Y_{t-1}, y_t) \\ &= \max_{S_{t-1}} P(y_t | S_{t-1}, s_t = e_i, Y_{t-1}) P(s_t = e_i | S_{t-1}, Y_{t-1}) P(S_{t-1}, Y_{t-1}) \\ &\approx \max_j \{ P(y_t | s_t = e_i, s_{t-1} = e_j, S_{t-2}^*(j), Y_{t-1}) P(s_t = e_i | s_{t-1} = e_j) \\ &\quad \max_{S_{t-2}} P(S_{t-2}, s_{t-1} = e_j, Y_{t-1}) \} \\ &= \max_j \{ J_{t|t-1,i,j} J_{t-1,j} \} \end{aligned} \quad (\text{Eq. 6})$$

- 10 where we denote

$$J_{t|t-1,i,j} = P(y_t | s_t = e_i, s_{t-1} = e_j, S_{t-2}^*(j), Y_{t-1}) P(s_t = e_i | s_{t-1} = e_j) \quad (\text{Eq. 7})$$

as the “transition probability” from state  $j$  at time  $t-1$  to state  $i$  at time  $t$ . Since this analysis takes place relative to time  $t$ , we refer to  $s_t$  as the “switching state,” and to  $s_{t-1}$

as the “previous switching state.”  $S_{t-2}^*(i)$  is the “best” switching sequence up to time  $t-1$  when SLDS is in state  $i$  at time  $t-1$ :

$$S_{t-2}^*(i) = \arg \max_{S_{t-2}} J_{t-1,i} \quad (\text{Eq. 8}) .$$

Hence, the switching sequence posterior at time  $t$  can be recursively computed from the same at time  $t-1$ . The two scaling components in  $J_{t|t-1,i,j}$  are the likelihood associated with the transition  $j \rightarrow i$  from  $t-1$  to  $t$ , and the probability of discrete SLDS switching from  $j$  to  $i$ .

To find the likelihood term, note that concurrently with the recursion of Equation 6, for each pair of consecutive switching states  $j, i$  at times  $t-1, t$ , one can obtain the following statistics using the Kalman filter:

$$\begin{aligned} \hat{X}_{t|t,i} &\triangleq \langle x_t | Y_t, s_t = e_i \rangle \\ \Sigma_{t|t,i} &\triangleq \langle (x_t - \hat{x}_{t|t,i})(x_t - \hat{x}_{t|t,i})' | Y_t, s_t = e_i \rangle \\ \hat{x}_{t|t-1,i,j} &\triangleq \langle x_t | Y_{t-1}, s_t = e_i, s_{t-1} = e_j \rangle \\ \Sigma_{t|t-1,i,j} &\triangleq \langle (x_t - \hat{x}_{t|t,i,j})(x_t - \hat{x}_{t|t,i,j})' | Y_t, s_t = e_i, s_{t-1} = e_j \rangle \\ \hat{x}_{t|t,i,j} &\triangleq \langle x_t | Y_t, s_t = e_i, s_{t-1} = e_j \rangle \\ \Sigma_{t|t,i,j} &\triangleq \langle (x_t - \hat{x}_{t|t,i,j})(x_t - \hat{x}_{t|t,i,j})' | Y_t, s_t = e_i, s_{t-1} = e_j \rangle \end{aligned}$$

where  $\hat{x}_{t|t,i}$  is the “best” filtered LDS state estimate at  $t$  when the switch is in state  $i$  at time  $t$  and a sequence of  $t$  measurements,  $Y_t$ , has been processed;  $\hat{x}_{t|t-1,i,j}$  and  $\hat{x}_{t|t,i,j}$  are the one-step predicted LDS state and the “best” filtered state estimates at time  $t$ , respectively, given that the switch is in state  $i$  at time  $t$  and in state  $j$  at time  $t-1$  and only

$t-1$  measurements are known. The two sets  $\{\hat{x}_{t|t-1,i,j}\}$  and  $\{\hat{x}_{t|t,i,j}\}$ , where  $i$  and  $j$  take on all possible values, are examples of “sets of continuous state estimates.” The set  $\{\hat{x}_{t|t-1,i,j}\}$  is obtained through “Viterbi prediction.” Similar definitions are easily obtained for filtered and predicted state variance estimates,  $\Sigma_{t|i}$  and  $\Sigma_{t|t-1,i,j}$  respectively. For a given switch state transition  $j \rightarrow i$  it is now easy to establish relationship between the filtered and the predicted estimates. From the theory of Kalman estimation, it follows that for transition  $j \rightarrow i$  the following time updates hold:

$$\hat{x}_{t|t,i,j} = A_i \hat{x}_{t-1|t-1,j} \quad (\text{Eq. 9})$$

$$\Sigma_{t|t,i,j} = A_i \Sigma_{t-1|t-1,j} A_i' + Q_i \quad (\text{Eq. 10})$$

Given a new observation  $y_t$  at time  $t$ , each of these predicted estimates can now be filtered using a Kalman “measurement update” framework. For instance, the state estimate measurement update equation yields

$$\hat{x}_{t|t,i,j} = \hat{x}_{t|t-1,i,j} + K_{i,j} (y_t - C \hat{x}_{t|t-1,i,j}). \quad (\text{Eq. 11})$$

where  $K_{i,j}$  is the Kalman gain matrix associated with the transition  $j \rightarrow i$ .

Appropriate equations can be obtained that link  $\Sigma_{t|t-1,i,j}$  and  $\Sigma_{t|t,i,j}$ . The likelihood term can then be easily computed as the probability of innovation  $y_t - C \hat{x}_{t|t-1,i,j}$  of  $j \rightarrow i$  transition,

$$y_t | s_t = e_i, s_{t-1} = e_j, S_{t-2}^*(j) \sim N(y_t; C \hat{x}_{t|t-1,i,j}, C \Sigma_{t|t-1,i,j} C' + R) \quad (\text{Eq. 12})$$

Obviously, for every current switching state  $i$  there are  $S$  possible previous switching states. To maximize the overall probability at every time step and for every switching state, one “best” previous state  $j$  is selected:

$$\psi_{t-1,i} = \arg \max_j \{J_{t|t-1,i,j} J_{t-1,j}\} \quad (\text{Eq. 13})$$

Since the best state is selected based on the continuous state predictions from the previous time step,  $\psi_{t-1,i}$  is referred to as the “optimal prior switching state.” The index of this state is kept in the state transition record entry  $\psi_{t-1,i}$ . Consequently, we now obtain a set of  $S$  best filtered LDS states and variances at time  $t$ :

$$5 \quad \hat{x}_{t|i,j} = \hat{x}_{t|i,\psi_{t-1,i}} \text{ and } \Sigma_{t|i,j} = \Sigma_{t|i,\psi_{t-1,i}}.$$

Once all  $T$  observations  $Y_{T-1}$  have been fused to decode the “best” switching state sequence, one uses the index of the best final state,  $i_{T-1}^* = \arg \max_j J_{T-1,i}$ , and

then traces back through the state transition record  $\psi_{t-1,i}$ , setting  $i_t^* = \psi_{t,i_{t+1}^*}$ . The

switching model's sufficient statistics are now simply  $\langle s_t \rangle = e_{i_t^*}$  and  $\langle s_t s'_{t-1} \rangle = e_{i_t^*} e'_{i_{t-1}^*}$ .

- 10 Given the “best” switching state sequence, the sufficient LDS statistics can be easily obtained using Rauch-Tung-Streiber (RTS) smoothing. Smoothing is described in Anderson et al, “Optimal Filtering,” Prentice-Hall, Inc., Englewood Cliffs, NJ, 1979. For example,

$$\langle x_t, s_t(i) \rangle = \begin{cases} \hat{x}_{t|T-1,i^*} & i = i_t^* \\ 0 & \text{otherwise} \end{cases}$$

- 15 for  $i = 0, \dots, S-1$ .

Figs. 4A and 4B comprise a flowchart that summarizes an embodiment of the present invention employing the Viterbi inference algorithm for SLDSs, as described above. The steps are as follows:

- 20 Initialize LDS state estimates  $\hat{x}_{0-1,i}$  and  $\Sigma_{0-1,i}$ ; (Step 102)  
 Initialize  $J_{0,i}$ . (Step 102)  
     for  $i = 1:T-1$  (Steps 104, 122)  
         for  $j = 1:S$  (Steps 106, 120)  
             for  $j = 1:S$  (Steps 108, 114)  
                 Predict and filter LDS state estimates  
 25  $\hat{x}_{t|i,j}$  and  $\Sigma_{t|i,j}$  (Step 110)  
                 Find  $j \rightarrow i$  “transition probability”  $J_{t-1,i,j}$



end (Step 112)  
 Find best transition  $J_{t,i}$ , into state  $i$ ; (Step 116)  
 Update sequence probabilities  $J_{t,i}$  and LDS state  
 estimates  $\hat{x}_{t|i}$  and  $\Sigma_{t|i}$  (Step 118)  
 5 end  
 Find “best” final switching state  $i_{T-1}^*$  (Step 124)  
 Backtrack to find “best” switching state sequence  $i_t^*$  (Step 126)  
 Find DBN's sufficient statistics. (Step 128)

#### APPROXIMATE VARIATIONAL INFERENCE EMBODIMENT

10 Fig. 5 is a block diagram for an embodiment of the dynamics learning method based on approximate variational inference. Generally, reference numbers 40 - 46 correspond to reference numbers 30 - 36 of Fig. 3, the approximate Viterbi inference block 34 of Fig. 3 being replaced by the approximate variational inference block 44.

At step 42, the switching dynamics of one or more SLDS motion models are  
 15 learned from a corpus of state space motion examples 40. Parameters 46 of each SLDS are re-estimated, at step 44, iteratively so as to minimize the modeling cost of current state sequence estimates that are obtained using the approximate variational inference procedure. Approximate variational inference is developed as an alternative to computationally expensive exact state sequence estimation.

20 A general structured variational inference technique for Bayesian networks is described in Jordan et al., “An Introduction to Variational Methods For Graphical Models,” Learning In Graphical Models, Kluwer Academic Publishers, 1998. They consider a parameterized distribution  $Q$  which is in some sense close to the desired conditional distribution  $P$ , but which is easier to compute.  $Q$  can then be employed as  
 25 an approximation of  $P$ ,

$$P(X_T, S_T | Y_T) \approx Q(X_T, S_T | Y_T)$$

Namely, for a given set of observations  $Y_T$ , a distribution  $Q(X_T, S_T | \eta, Y_T)$  with an additional set of *variational parameters*  $h$  is defined such that Kullback-Leibler divergence between  $Q(X_T, S_T | \eta, Y_T)$  and  $P(X_T, S_T | Y_T)$  is minimized with respect to  $h$ :

$$\eta^* = \arg \min_{\eta} \sum_{S_T} \int_{X_T} Q(X_T, S_T | \eta, Y_T) \log \frac{P(X_T, S_T | Y_T)}{Q(X_T, S_T | \eta, Y_T)}.$$

The dependency structure of  $Q$  is chosen such that it closely resembles the dependency structure of the original distribution  $P$ . However, unlike  $P$ , the dependency structure of  $Q$  must allow a computationally efficient inference. In our case, we define  $Q$  by decoupling the switching and LDS portions of SLDS as shown in Fig. 6.

Fig. 6 illustrates the factorization of the original SLDS. The two subgraphs of the original network are a Hidden Markov Model (HMM)  $Q_S$  50 with variational parameters  $\{q_0, \dots, q_{T-1}\}$ , and a time-varying LDS.  $Q_X$  52 with variational parameters  $\{\hat{x}_0, \hat{A}_0, \dots, \hat{A}_{T-1}, \hat{Q}_0, \dots, \hat{Q}_{T-1}\}$ . More precisely, the Hamiltonian of the approximating dependency graph is defined as:

$$\begin{aligned} H_Q(X_T, S_T, Y_T) = & \frac{1}{2} \sum_{t=1}^{T-1} (x_t - \hat{A}_t x_{t-1})' \hat{Q}_t^{-1} (x_t - \hat{A}_t x_{t-1}) + \frac{1}{2} \log |\hat{Q}_t| + \\ & \frac{1}{2} (x_0 - \hat{x}_0)' \hat{Q}_0^{-1} (x_0 - \hat{x}_0) + \frac{1}{2} \log |\hat{Q}_0| + \frac{NT}{2} \log 2\pi + \\ & \frac{1}{2} \sum_{t=0}^{T-1} (y_t - Cx_t)' R^{-1} (y_t - Cx_t) + \frac{T}{2} \log |R| + \frac{MT}{2} \log 2\pi \\ & + \sum_{t=1}^{T-1} s'_t (-\log \Pi) s_{t-1} + s'_0 (-\log \pi_0) + \sum_{t=1}^{T-1} s'_t (-\log q_t). \end{aligned} \quad (\text{Eq. 14})$$

The two subgraphs 50, 52 are “decoupled,” thus allowing for independent inference,  $Q(X_T, S_T | \eta, Y_T) = Q(X_T | \eta, Y_T) Q_S(S_T | \eta)$ . This is also reflected in the sufficient statistics of the posterior defined by the approximating network, e.g.,

$$\langle x_i, x'_i s_i \rangle = \langle x_i, x'_i \rangle \langle s_i \rangle.$$

- 5 The optimal values of the variational parameters  $h$  are obtained by setting the derivative of the KL-divergence with respect to  $h$  to zero. We can then arrive at the following optimal variational parameters:

$$\begin{aligned} \hat{Q}_{T-1}^{-1} &= \sum_{i=0}^{S-1} Q_i^{-1} \langle s_i(i) \rangle \\ \hat{Q}_t^{-1} &= \sum_{i=0}^{S-1} Q_i^{-1} \langle s_i(i) \rangle + \sum_{i=0}^{S-1} A'_i Q_i^{-1} \langle s_{i+1}(i) \rangle - \hat{A}'_{t+1} \hat{Q}_{t+1}^{-1} \hat{A}_{t+1}, 0 < t < T-1 \\ \hat{Q}_0^{-1} &= \sum_{i=0}^{S-1} Q_{0,i}^{-1} \langle s_0(i) \rangle + \sum_{i=0}^{S-1} A'_i Q_i^{-1} A_i \langle s_1(i) \rangle - \hat{A}'_1 \hat{Q}_1^{-1} \hat{A}_1 \\ \hat{A}_t &= \hat{Q}_t \sum_{i=0}^{S-1} Q_i^{-1} A_i \langle s_i(i) \rangle \\ \hat{x}_0 &= \hat{Q}_0 \sum_{i=0}^{S-1} Q_{0,i}^{-1} x_{0,i} \langle s_0(i) \rangle \end{aligned} \quad (\text{Eq. 15})$$

$$\begin{aligned} \log q_0(i) &= -\frac{1}{2} \langle (x_0 - x_{0,i})' \hat{Q}_{0,i}^{-1} (x_0 - x_{0,i}) \rangle - \frac{1}{2} \log |\hat{Q}_{0,i}| \\ \log q_t(i) &= -\frac{1}{2} \langle (x_t - A_t x_{t-1})' \hat{Q}_t^{-1} (x_t - A_t x_{t-1}) \rangle - \frac{1}{2} \log |\hat{Q}_t|, t > 0 \end{aligned} \quad (\text{Eq. 16})$$

- 10 To obtain the expectation terms  $\langle s_t \rangle = \Pr(s_t | q_0, \dots, q_{T-1})$ , we use the inference in the HMM with output “probabilities”  $q_t$ , as described in Rabiner and Juang, “Fundamentals of Speech Recognition,” Prentice Hall, 1993. Similarly, to obtain

$\langle x_t \rangle = E[x_t | Y_T]$ , LDS inference is performed in the decoupled time-varying LDS via RTS smoothing. Since  $\hat{A}_t, \hat{Q}_t$  in the decoupled LDS  $Q_X$  depend on  $\langle s_t \rangle$  from the decoupled HMM  $Q_S$ , and  $q_t$  depends on  $\langle x_t \rangle, \langle x_t x'_t \rangle, \langle x_t x'_{t-1} \rangle$  from the decoupled LDS, Equations 15 and 16 together with the inference solutions in the decoupled models form a set of fixed-point equations. Solution of this fixed-point set yields a tractable approximation to the intractable inference of the original fully coupled SLDS.

Fig. 7 is a flowchart summarizing the variational inference algorithm for fully coupled SLDSs, corresponding to the steps below:

```

error = ∞;                                     (Step 152)
10  Initialize  $\langle s_t \rangle$ ;                         (Step 152)
    while (error > maxError) {                  (Step 154)
        Find  $\hat{Q}_t, \hat{A}_t, \hat{x}_0$  from  $\langle s_t \rangle$  using Equations 11; (Step 156)
        Estimate  $\langle x_t \rangle, \langle x_t x'_t \rangle$  and  $\langle x_t x'_{t-1} \rangle$  from  $Y_T$  using time-varying LDS
            inference;                             (Step 158)
15  Find  $q_t$  from  $\langle x_t \rangle, \langle x_t x'_t \rangle$  and  $\langle x_t x'_{t-1} \rangle$  using Equations 12;
                                            (Step 160)
        Estimate  $\langle s_t \rangle$  from  $q_t$  using HMM inference.         (Step 162)
        Update approximation error (KL divergence);             (Step 164)
    }

```

20 Variational parameters in Equations 15 and 16 have an intuitive interpretation. Variational parameters of the decoupled LDS  $\hat{Q}_t$  and  $\hat{A}_t$  in Equation 15 define a best unimodal (non-switching) representation of the corresponding switching system and are, approximately, averages of the original parameters weighted by a best estimates of the switching states  $P(s)$ . Variational parameters of the decoupled HMM  $\log q_t(i)$  in

25 Equation 16 measure the agreement of each individual LDS with the data.

## GENERAL PSEUDO BAYESIAN INFERENCE EMBODIMENT

The Generalized Pseudo Bayesian (GPB) approximation scheme first introduced in Bar-Shalom et al., "Estimation and Tracking: Principles, Techniques, and Software," Artech House, Inc. 1993 and in Kim, "Dynamic Linear Models With Markov-Switching," Journal of Econometrics, volume 60, pages 1-22, 1994, is based on the idea of "collapsing", i.e., representing a mixture of  $M^t$  Gaussians with a mixture of  $M^r$  Gaussians, where  $r < t$ . A detailed review of a family of similar schemes appears in Kevin P. Murphy, "Switching Kalman Filters," Technical Report 98-10, Compaq Cambridge Research Lab, 1998. We develop a SLDS inference scheme jointly in the switching and linear dynamic system states, derived from filtering GPB2 approach of Bar-Shalom et al. The algorithm maintains a mixture of  $M^2$  Gaussians over all times.

GPB2 is closely related to the Viterbi approximation described previously. It differs in that instead of picking the most likely previous switching state  $j$  at every time step  $t$  and switching state  $i$ , we collapse the  $M$  Gaussians (one for each possible value of  $j$ ) into a single Gaussian.

Consider the filtered and predicted means  $\hat{x}_{t|i,j}$  and  $\hat{x}_{t-1|i,j}$ , and their associated covariances, which were defined previously with respect to the approximate Viterbi embodiment. Assume that for each switching state  $i$  and pairs of states  $(i,j)$  the following distributions are defined at each time step:

$$\begin{aligned} \Pr(s_t = i | Y_t) \\ \Pr(s_t = i, s_{t-1} = j | Y_t) \end{aligned}$$

Regular Kalman filtering update can be used to fuse the new measurement  $y_t$  and obtain  $S^2$  new SLDS states at  $t$  for each  $S$  states at time  $t-1$ , in a similar fashion to the Viterbi approximation embodiment discussed previously.

Unlike Viterbi approximation which picks one best switching transition for each state  $i$  at time  $t$ , GPB2 "averages" over  $S$  possible transitions from  $t-1$ . Namely, it is easy to see that

$$\Pr(s_t = i, s_{t-1} = j | Y_t) \approx \Pr(y_t | \hat{x}_{t,i,j}) \Pi(i, j) \Pr(s_{t-1} = j | Y_{t-1}) \text{ .}$$

From there it follows immediately that the current distribution over the switching states is  $\Pr(s_t = i | Y_t) = \sum_j \Pr(s_t = i, s_{t-1} = j | Y_t)$  and that each previous state  $j$  now has

the following posterior

$$5 \quad \Pr(s_{t-1} = j | s_t = i, Y_t) = \frac{\Pr(s_t = i, s_{t-1} = j | Y_t)}{\Pr(s_t = i | Y_t)}.$$

This posterior is important because it allows one to “collapse” or “average” the S transitions into each state  $i$  into one average state, e.g.

$$\hat{x}_{t|i} = \sum_j \hat{x}_{t|i,j} \Pr(s_{t-1} = j | s_t = i, Y_t).$$

Analogous expressions can be obtained for the variances  $S_{t|t,i}$  and  $S_{t-1|t,i}$

Smoothing in GPB2 is a more involved process that includes several additional approximations. Details can be found in Kevin P. Murphy, "Switching Kalman Filters," Technical Report 98-10, Compaq Cambridge Research Lab, 1998. First an assumption is made that decouples the switching model from the LDS when smoothing the switching states. Smoothed switching states are obtained directly from  $\Pr(s_t|Y)$  estimates, namely  $\Pr(s_t = i \mid s_{t+1} = k, Y_T) \approx \Pr(s_t = i \mid s_{t+1} = k, Y_t)$ . Additionally, it is assumed that  $\hat{x}_{t+1|T,i,k} \approx \hat{x}_{t+1|T,k}$ . These two assumptions lead to a set of smoothing equations for each transition (i,k) from t to t+1 that obey an RTS smoother, followed by collapsing similar to the filtering step.

Figs. 8A and 8B comprise a flowchart illustrating the steps employed by a  
20 GPB2 embodiment, as summarized by the following steps:

Initialize LDS state estimates  $\hat{x}_{0|-1,i}$  and  $\Sigma_{0|-1,i}$ ; (Step 202)

```

Initialize  $\Pr(s_0 = i) = p(i)$ , for  $i=0, \dots, S-1$ ; (Step 202)
for  $t = 1:T-1$  (Steps 204, 222)
    for  $i = 1:S$  (Steps 206, 220)
        for  $j = 1:S$  (Steps 208, 216)
            Predict and filter LDS state estimates  $\hat{x}_{t|i,j}, \Sigma_{t|i,j}$ ; (Step 210)

            Find switching state distributions
             $\Pr(s_t = i|Y_t), \Pr(s_{t-1} = j|s_t = i, Y_t)$ ; (Step 212)
            Collapse  $\hat{x}_{t|i,j}, \Sigma_{t|i,j}$  to  $\hat{x}_{t|i}, \Sigma_{t|i}$ ; (Step 214)

10         end
            Collapse  $\hat{x}_{t|i}, \Sigma_{t|i}$  to  $\hat{x}_{t|t}$  and  $\Sigma_{t|t}$ . (Step 218)

        end
    end
    Do GPB2 smoothing; (Step 224)
15    Find sufficient statistics. (Step 226)

```

The inference process of the GPB2 embodiment is clearly more involved than those of the Viterbi or the variational approximation embodiments. Unlike Viterbi, GPB2 provides soft estimates of switching states at each time  $t$ . Like Viterbi, GPB2 is a local approximation scheme and as such does not guarantee global optimality inherent in the variational approximation. However, Xavier Boyen and Daphne Koller, “Tractable inference for complex stochastic processes,” *Uncertainty in Artificial Intelligence*, pages 33-42, Madison, WI, 1998, provides complex conditions for a similar type of approximation in general DBNs that lead to globally optimal solutions.

## LEARNING OF SLDS PARAMETERS

Learning in SLDSs can be formulated as the problem of maximum likelihood learning in general Bayesian networks. Hence, a generalized Expectation-

Maximization (EM) algorithm can be used to find optimal values of SLDS parameters  $\{A_0, \dots, A_{s-1}, C, Q_0, \dots, Q_{s-1}, R, \Pi, \pi_0\}$ . A description of generalized EM can be found

in, for example, Neal et al., “A new view of the EM algorithm that justifies incremental and other variants,” in the collection “Learning in graphical models,” (M. Jordan,

5 editor), pages 355-368. MIT Press, 1999.

The EM algorithm consists of two steps, E and M, which are interleaved in an iterative fashion until convergence. The essential step is the expectation (E) step. This step is also known as the inference problem. The inference problem was considered in the two preferred embodiments of the method.

Given the sufficient statistics from the inference phase, it is easy to obtain parameter update equations for the maximization (M) step of the EM algorithm.

Updated values of the model parameters are easily shown to be

$$\begin{aligned}\hat{A}_i &= \left( \sum_{t=1}^{T-1} \langle x_t x'_{t-1} s_t(i) \rangle \right) \left( \sum_{t=1}^{T-1} \langle x_{t-1} x'_{t-1} s_t(i) \rangle \right)^{-1} \\ \hat{Q}_i &= \left( \sum_{t=1}^{T-1} \langle x_t x'_i s_t(i) \rangle - \hat{A}_i \langle x_{t-1} x'_i s_t(i) \rangle \right) \left( \sum_{t=1}^{T-1} \langle s_t(i) \rangle \right)^{-1} \\ \hat{C} &= \left( \sum_{t=0}^{T-1} y_t \langle x'_t \rangle \right) \left( \sum_{t=0}^{T-1} \langle x_t x' \rangle \right)^{-1} \\ \hat{R} &= \frac{1}{T} \sum_{t=0}^{T-1} (y_t y'_t - \hat{C} \langle x_t \rangle y'_t) \\ \hat{\Pi} &= \left( \sum_{t=1}^{T-1} \langle s_t s'_{t-1} \rangle \right) \text{diag} \left( \sum_{t=1}^{T-1} \langle s_t \rangle \right)^{-1} \\ \hat{\pi}_0 &= \langle s_0 \rangle.\end{aligned}$$



The operator  $\langle \cdot \rangle$  denotes conditional expectation with respect to the posterior distribution, e.g.  $\langle x_t \rangle = \sum_S \int_X x_t P(X, S | Y)$ .

All the variable statistics are evaluated before updating any parameters. Notice that the above equations represent a generalization of the parameter update equations of classical (non-switching) LDS models.

The coupled E and M steps are the key components of the SLDS learning method. While the M step is the same for all preferred embodiments of the method, the E-step will vary with the approximate inference method.

#### APPLICATIONS OF SLDS

We applied our SLDS framework to the analysis of two categories of fronto-parallel motion: walking and jogging. Fronto-parallel motions exhibit interesting dynamics and are free from the difficulties of 3-D reconstruction. Experiments can be conducted easily using a single video source, while self-occlusions and cluttered backgrounds make the tracking problem non-trivial.

The kinematics of the figure are represented by a 2-D Scaled Prismatic Model (SPM). This model is described in Morris et al., "Singularity analysis for articulated object tracking," Proceedings of Computer Vision and Pattern Recognition, pages 289-296, Santa Barbara, CA, June 23-25, 1998. The SPM lies in the image plane, with each link having one degree of freedom (DOF) in rotation and another DOF in length. A chain of SPM transforms can model the image displacement and foreshortening effects produced by 3-D rigid links. The appearance of each link in the image is described by a template of pixels which is manually initialized and deformed by the link's DOFs.

In our figure tracking experiments, we analyzed the motion of the legs, torso, and head, while ignoring the arms. Our kinematic model had eight DOFs, corresponding to rotations at the knees, hip and neck.

Learning

The first task we addressed was learning an SLDS model for walking and running. The training set consisted of eighteen sequences of six individuals jogging and walking at a moderate pace. Each sequence was approximately fifty frames in duration. The training data consisted of the joint angle states of the SPM in each image frame, which were obtained manually.

The two motion types were each modeled as multi-state SLDSs and then combined into a single complex SLDS. The measurement matrix in all cases was assumed to be identity,  $C = I$ . Initial state segmentation within each motion type was obtained using unsupervised clustering in a state space of some simple dynamics model, e.g., a constant velocity model. Parameters of the model ( $A, Q, R, x_0, P, \pi_0$ ) were then reestimated using the EM-learning framework with approximate Viterbi inference. This yielded refined segmentation of switching states within each of the models.

Fig. 9 illustrates learned segmentation of a “jog” motion sequence. A two-state SLDS model was learned from a set of exemplary “jog” motion measurement sequences, an example of which is shown in the bottom graph. The top graph depicts decoded switching states ( $\langle s_t \rangle$ ) inferred from the measurement sequence  $y_t$ , shown in the bottom graph, using the learned “jog” SLDS model.

### Classification

The task of classification is to segment a state space trajectory into a sequence of motion regimes, according to learned models. For instance, in gesture recognition, one can automatically label portions of a long hand trajectory as some predefined, meaningful gestures. The SLDS framework is particularly suitable for motion trajectory classification.

Fig. 10 illustrates the classification of state space trajectories. Learned SLDS model parameters are used in approximate Viterbi inference to decode a best sequence of models corresponding to the state space trajectory to be classified.

The classification framework follows directly from the framework for approximate Viterbi inference in SLDSs, described previously. Namely, the

approximate Viterbi inference 74 yields a sequence of switching states  $S_T$  (regime indexes) 70 that best describes the observed state space trajectory 70, assuming some current SLDS model parameters. If those model parameters are learned on a corpus of representative motion data, applying the approximate Viterbi inference on a state  
 5 trajectory from the same family of motions will then result in its segmentation into the learned motion regimes.

Additional “constraints” 72 can be imposed on classification. Such constraints 72 can model expert knowledge about the domain of trajectories which are to be classified, such as domain grammars, which may not have been present during SLDS  
 10 model learning. For instance, we may know that out of  $N$  available SLDS motion models, only  $M < N$  are present in the unclassified data. We may also know that those  $M$  models can only occur in some particular, e.g., deterministically or stochastically, known order. These classification constraints can then be superimposed on the SLDS motion model parameters in the approximate Viterbi inference to force the  
 15 classification to adhere to them. Hence, a number of natural language modeling techniques from human speech recognition, for example, as discussed in Jelinek, “Statistical methods for speech recognition,” MIT Press, 1998, can be mapped directly into the classification SLDS domain.

Classification can also be performed using variational and GPB2 inference  
 20 techniques. Unlike with Viterbi inference, these embodiments yield a “soft” classification, i.e., each switching state at every time instance can be active with some potentially non-zero probability. A soft switching state at time  $t$  is  $\sum_{i=0}^{S-1} i \langle s_t(i) \rangle$ .

To explore the classification ability of our learned model, we next considered segmentation of sequences of complex motion, i.e., motion consisting of alternations of  
 25 “jog” and “walk.” Test sequences were constructed by concatenating in random order randomly selected and noise-corrupted training sequences. Transitions between sequences were smoothed using B-spline smoothing. Identification of two motion

“regimes” was conducted using the proposed framework in addition to a simple HMM-based segmentation. Multiple choices of linear dynamic system and “jog” / “walk” switching orders were also compared.

Because “jog” and “walk” SLDS models were learned individually, whereas the  
 5 test data contained mixed “jog” and “walk” regimes, it was necessary to impose classification constraints to combine the two models into a single “jog + walk” model. Since no preference was known a priori for either of the two regimes, the two were assumed equally likely but complementary, and their individual two-state switching models  $P_{\text{jog}}$  and  $P_{\text{walk}}$  were concatenated into a single four-state switching model  $P_{\text{jog+walk}}$ .  
 10 Each state of this new model simply carried over the LDS parameters of the individual “jog” and “walk” models.

Estimates of “best” switching states  $\langle s_t \rangle$  indicated which of the two models can be considered to be driving the corresponding motion segment.

Fig. 11 illustrates an example of segmentation, depicting true switching state  
 15 sequence (sequence of jog-walk motions) in the top graph, followed by HMM, Viterbi, GPB2, and variational state estimates using one switching state per motion type models, first order SLDS model.

Fig. 11 contains several graphs which illustrate the impact of different learned models and inference methods on classification of “jog” and “walk” motion sequences,  
 20 where learned “jog” and “walk” motion models have two switching states each. Continuous system states of the SLDS model contain information about angle of the human figure joints. The top graph 80 depicts correct classification of a motion sequence containing “jog” and “walk” motions (measurement sequence is not shown). The remaining graphs 82 - 88 show inferred classifications using, respectively from top  
 25 to bottom, SLDS model with Viterbi inference, SLDS model with GPB2 inference, SLDS model with variational inference, and HMM model.

Similar results were obtained with four switching states each and second order SLDS.

Classification experiments on a set of 20 test sequences gave an error rate of 2.9% over a total of 8312 classified data points, where classification error was defined as the difference between inferred segmentation (the estimated switching states) and true segmentation (true switching states) accumulated over all sequences,

$$5 \quad error = \sum_{t=0}^{T-1} |\langle s_t \rangle - s_{true,t}|.$$

### Tracking

The SLDS learning framework can improve the tracking of target objects with complex dynamical behavior, responsive to a sequence of measurements. A particular embodiment of the present invention tracks the motion of the human figure, using  
 10 frames from a video sequence. Tracking systems generally have three identifiable steps: state prediction, measurement processing and posterior update. These steps are exemplified by the well-known Kalman Filter, described in Anderson and Moore, "Optimal Filtering," Prentice Hall, 1979.

Fig. 12 is a standard block diagram for a Kalman Filter. A state prediction  
 15 module 500 takes as input the state estimate from the previous time instant,  $\hat{x}_{t-1|t-1}$ .

The output of the prediction module 500 is the predicted state  $\hat{x}_{t|t-1}$ . The measurement processing module 501 takes the predicted state, generates a corresponding predicted measurement, and combines it with the actual measurement  $y_t$  to form the "innovation"  $z_t$ . For an image, for example, states might be parameters such as angles, lengths and  
 20 positions of objects or features, while the measurements might be the actual pixels. The predicted measurements might then be the predictions of pixel values and/or pixel locations. The innovation  $z_t$  measures the difference between the predicted and actual measurement data.

The innovation  $z_t$  is passed to the posterior update module 502 along with the prediction. They are fused using the Kalman gain matrix to form the posterior estimate  $\hat{x}_{t|t}$ .

The delay element 503 indicates the temporal nature of the tracking process. It models the fact that the posterior estimate for the current frame becomes the input for the filter in the next frame.

Fig. 13 illustrates the operation of the prediction module 500 for the specific case of figure tracking. The dashed line 510 shows the position of a skeletal representation of the human figure at time  $t - 1$ . The predicted position of the figure at time  $t$  is shown as a solid line 511. The actual position of the figure in the video frame at time  $t$  is shown as a dotted line 512. The unknown state  $x_t$  represents the actual skeletal position for the sketched figure.

We now describe the process of computing the innovation  $z_t$ , which is the task of the measurement processing module 501, in the specific case of figure tracking using template features. A template is a region of pixels, often rectangular in shape. Tracking using template features is described in detail in James M. Rehg and Andrew P. Witkin, "Visual Tracking with Deformation Models," Proceedings of IEEE Conference on Robotics and Automation, April 1991, Sacramento CA, pages 844-850. In the figure tracking application, a pixel template is associated with each part of the figure model, describing its image appearance. The pixel template is an example of an "image feature model" which describes the measurements provided by the image.

For example, two templates can be used to describe the right arm, one for the lower and one for the upper arm. These templates, along with those for the left arm and the torso, describe the appearance of the subject's shirt. These templates can be  
25 initialized, for example, by capturing pixels from the first frame of the video sequence in which each part is visible. The use of pixel templates in figure tracking is described in detail in Daniel D. Morris and James M. Rehg, "Singularity Analysis for Articulated

Object Tracking,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 1998, Santa Barbara CA, pages 289-296.

Given a set of initialized figure templates, the innovation is the difference between the template pixel values and the pixel values in the input video frame which correspond to the predicted template position. We can define the innovation function that gives the pixel difference as a function of the state vector  $x$  and pixel index  $k$ :

$$z(x, k) = I_t(\text{pos}(x, k)) - T(k) \quad (\text{Eq. 17})$$

We can then write the innovation at time  $t$  as

$$z_t(k) = z(\hat{x}_{t|t-1}, k) \quad (\text{Eq. 18})$$

Equation 18 defines a vector of pixel differences,  $z_t$ , formed by subtracting pixels in the template model  $T$  from a region of pixels in the input video frame  $I_t$  under the action of the figure state. In order to represent images as vectors, we require that each template pixel in the figure model be assigned a unique index  $k$ . This can be easily accomplished by examining the templates in a fixed order and scanning the pixels in each template in raster order. Thus  $z_t(k)$  represents the innovation for the  $k$ th template pixel, whose intensity value is given by  $T(k)$ . Note that the contents of the template  $T(k)$  could be, for example, gray scale pixel values, color pixel values in RGB or YUV, Laplacian filtered pixels, or in general any function applied to a region of pixels. We use  $z(x)$  to denote the vector of pixel differences described by Equation (e1). Note that  $z_t$  does not define an innovation process in the strict sense of the Kalman filter, since the underlying system is nonlinear and non-Gaussian.

The deformation function  $\text{pos}(x, k)$  in Equation 17 gives the position of the  $k$ th template pixel, with respect to the input video frame, as a function of the model state  $x$ . Given a state prediction  $\hat{x}_{t|t-1}$ , the transform  $\text{pos}(\hat{x}_{t|t-1}, *)$  maps the template pixels into the current image, thereby selecting a subset of pixel measurements. In the case of

figure tracking, the  $\text{pos}()$  function models the kinematics of the figure, which determine the relative motion of its parts.

Fig. 14 illustrates two templates 513 and 514 which model the right arm 525. Pixels that make up these templates 513, 514 are ordered, with pixels  $T_1$  through  $T_{100}$  belonging to the upper arm template 513, and pixels  $T_{101}$  through  $T_{200}$  belonging to the lower arm template 514. The mapping  $\text{pos}(x, k)$  is also illustrated. The location of each template 513, 514 under the transformation is shown as 513A, 514A respectively. The specific transformations of two representative pixel locations,  $T_{48}$  and  $T_{181}$ , are also illustrated.

Alternatively, the boundary of the figure in the image can be expressed as a collection of image contours whose shapes can be modeled for each part of the figure model. These contours can be measured in an input video frame and the innovation expressed as the distance in the image plane between the positions of predicted and measured contour locations. The use of contour features for tracking is described in detail in Demetri Terzopoulos and Richard Szeliski, "Tracking with Kalman Snakes," which appears in "Active Vision," edited by Andrew Blake and Alan Yuille, MIT Press, 1992, pages 3-20. In general, the tracking approach outlined above applies to any set of features which can be computed from a sequence of measurements.

In general, each frame in an image sequence generates a set of "image feature measurements," such as pixel values, intensity gradients, or edges. Tracking proceeds by fitting an "image feature model," such as a set of templates or contours, to the set of measurements in each frame.

A primary difficulty in visual tracking is the fact that the image feature measurements are related to the figure state only in a highly nonlinear way. A change in the state of the figure, such as raising an arm, induces a complex change in the pixel values produced in an image. The nonlinear function  $I(\text{pos}(x, *))$  models this effect.

The standard approach to addressing this nonlinearity is to use the Iterated Extended Kalman Filter (IEKF), which is described in Anderson and Moore, section 8.2. In this



approach, the nonlinear measurement model in Equation 17 is linearized around the state prediction  $\hat{x}_{t|t-1}$ .

The linearizing function can be defined as

$$M_t(x, k) = \nabla I_t(\text{pos}(x, k))' \frac{\partial \text{pos}}{\partial x}(x, k) \quad (\text{Eq. 19})$$

5 The first term on the right in Equation 19,  $\nabla I_t(\text{pos}(x, k))'$ , is the image gradient  $\nabla I_t$ , evaluated at the image position of template pixel  $k$ . The second term,

$\frac{\partial \text{pos}}{\partial x}(x, k)$ , is a  $2 \times N$  kinematic Jacobian matrix, where  $N$  is the number of states, or elements, in  $x$ . Each column of this Jacobian gives the displacement in the image at pixel position  $k$  due to a small change in one of the model states.  $M_t(x, k)$  is a  $1 \times N$  row vector. We denote by  $M_t(x)$  the  $M \times N$  matrix formed by stacking up these rows for each of the  $M$  pixels in the template model. The linearized measurement model can then be written:

$$\bar{C}_t = M_t(\hat{x}_{t|t-1}) \quad (\text{Eq. 20})$$

The standard Kalman filter equations can then be applied using the linearized measurement model from Equation 20 and the innovation from Equation 18. In particular, the posterior update equation (analogous to Equation 11 in the linear case) is:

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + L_t z_t \quad (\text{Eq. 21})$$

where  $L_t$  is the appropriate Kalman gain matrix formed using  $\bar{C}_t$ .

Fig. 15 illustrates a block diagram of the IEKF. In comparison to Fig. 12, the measurement processing 501 and posterior update 502 blocks within the dashed box 506 are now iterated  $P$  times with the same measurement  $y_t$ , for some predetermined number  $P$ , before the final output 507 is reported and a new measurement  $y_t$  is

introduced. This iteration produces a series of innovations  $z_t^n$  and posterior estimates  $\hat{x}_{t|t}^n$  which result from successive linearizations of the measurement model around the previous posterior estimate  $\hat{x}_{t|t}^{n-1}$ . The iterations are initialized by setting  $\hat{x}_{t|t}^1 = \hat{x}_{t|t-1}$ . Upon exiting, we set  $\hat{x}_{t|t} = \hat{x}_{t|t}^P$ .

5 In cases where the prediction is far from the correct state estimate, these iterations can improve accuracy. We denote as the “TEKF update module” 506 the subsystem of measurement processing 501 and posterior update 502 blocks, as well as delay 504, although in practice, delays 503 and 504 may be the same piece of hardware and/or software.

Clearly the quality of the IEKF solution depends upon the quality of the state prediction. The linearized approximation is only valid within a small range of state values around the current operating point, which is initialized by the prediction. More generally, there are likely to be many background regions in a given video frame that are similar in appearance to the figure. For example, a template that describes the appearance of a figure wearing a dark suit might match quite well to a shadow cast against a wall in the background. The innovation function defined in Equation 17 can be viewed as an objective function to be minimized with respect to  $x$ . It is clear that this function will in general have many local minima. The presence of local minima poses problems for any iterative estimation technique such as IEKF. An accurate prediction can reduce the likelihood of becoming trapped in a local minima by bringing the starting point for the iterations closer to the correct answer.

As alluded to previously, the dynamics of a complex moving object such as the human figure cannot be expressed using a simple linear dynamic model. Therefore, it is unlikely that the standard IEKF framework described above would be effective for figure tracking in video. A simple linear prediction of the figure's state would not provide sufficient accuracy in predicting figure motion.

10

defined by Equation 17. Note that the only difference in comparison to Equation 18 is the use of the SLDS prediction in mapping templates into the image plane.

15

9.

20

In the case of template features, the measurement probability can be written

$$\begin{aligned}
P(y_{t|t-1,i,j}) &= P(y_t | s_t = e_i, s_{t-1} = e_j, S_{t-2}^*(j)) \\
&\approx N(z_{t|t-1,i,j}; 0, \bar{C}_{t|t-1,i,j} \Sigma_{t|t-1,i,j} \bar{C}_{t|t-1,i,j}' + R)
\end{aligned} \tag{Eq. 23}$$

where  $\bar{C}_{t|t-1,i,j} = M_t(\hat{x}_{t|t-1,i,j})$ . The difference between Equations 23 and 12 is the use

of the linearized measurement model for template features. By changing the measurement probability appropriately, the SLDS framework above can be adopted to any tracking problem.

The output of the selector 511 is a set of S hypotheses corresponding to the most probable state predictions given the measurement data. These are input to the IEKF update block 506, which filters these predictions against the measurements to obtain a set of S posterior state estimates. This block was illustrated in Figure D1. This step applies the standard equations for the IEKF to features such as templates or contours, as described in Equations 17 through 21. The posterior estimates can be decoded and smoothed analogously to the case of Viterbi inference.

In computing the measurement probabilities, the selector 511 must make  $S^2$  comparisons between all of the model pixels and the input image. Depending upon the size of the target and the image, this may represent a large computation. This computation can be reduced by considering only the Markov process probabilities and not the measurement probabilities in computing the best S switching hypotheses. In this case, the best hypotheses are given by

$$i_t^* = \arg \min_j \{ -\log \Pi(i, j) + J_{t-1,j} \} \tag{Eq. 24}$$

This is a substantial reduction in computation, at the cost of a potential loss of accuracy in picking the best hypotheses.

Fig. 17 illustrates a second tracking embodiment, in which the order of the IEKF update module 506 and selector 511A is reversed. In this case, the set of  $S^2$  predictions are passed directly to the IEKF update module 506. The output of the

00T050-T045950

15  
JAN 8/29/00  
8/29/00

update module 506 is a set of  $S^2$  filtered estimates  $\hat{x}_{t|t,i,j}$ , each of which is the result of  $P$  iterations of the IEKF. As with the embodiment of Fig. 16, it is still necessary to reduce the total set of estimates to  $S$  hypotheses in order to control the complexity of the tracker.

5           The selector 511A chooses the most likely transition  $j \rightarrow i$  for each switching state based on the filtered estimates. This is analogous to the Viterbi inference case, which was described in the embodiment of Fig. 16 above. The key step is to compute the switching costs  $J_{t,i}$  defined in Equations 5 and 6. The difference in this case comes from the fact that the probability of the measurement is computed using the filtered  
10       posterior estimate rather than the prediction. This probability is given by:

$$P(y_{t|t,i,j}) = P(y_t | s_t = e_i, s_{t-1} = e_j, \hat{x}_{t|t,i,j}, S_{t-2}^*(j)) \\ \approx N(z(\hat{x}_{t|t,i,j}); 0, M_t(\hat{x}_{t|t,i,j}) \Sigma_{t|t,i,j} M_t(\hat{x}_{t|t,i,j})' + R)$$

The switching costs are then given by

$$J_{t,i} = \max_j \{J_{t|t,i,j} J_{t-1,j}\} \quad (\text{Eq. 25A})$$

where the “posterior transition probability” from state  $j$  at time  $t-1$  to state  $i$  at time  $t$  is  
15 written

$$J_{t|t,i,j} = P(y_{t|t,i,j})P(s_t = e_i | s_{t-1} = e_j) \quad (\text{Eq. 25B})$$

The selector 511A selects the transition  $j^* \rightarrow i$ , where

$j^* = \arg \max_j \{J_{t|t,i,j}, J_{t-1|j}\}$ . The state  $j^*$  is the “optimal posterior switching state,” and its value depends upon the posterior estimate for the continuous state at time

The potential advantage of this approach is that all of the  $S^2$  predictions are given an opportunity to filter the measurement data. Since the system is nonlinear, it is



5

$$\begin{aligned}\alpha_{i,j} &= p(s_t = e_i | s_{t-1} = e_j, S_{t-2}^*(j), Y_{t-1}) P(s_{t-1} = e_j, S_{t-2}^*(j) | Y_{t-1}) \\ &= P(s_t = e_i | s_{t-1} = e_j) \frac{P(s_{t-1} = e_j, S_{t-2}^*(j), Y_{t-1})}{P(Y_{t-1})}\end{aligned}$$

Applying Equation 1 to the first term and Equations 5 and 8 to the second term

10 All of the terms in the numerator of Equation 27 are directly available from the Viterbi inference method. The denominator is a constant normalizing factor which can be rewritten to yield

Equations 26 and 28 define a “Viterbi mixture density” from which additional starting points for search can be drawn. The steps for drawing R additional points are as follows:

For  $r = 1$  to  $R$

Select a kernel  $K_{i_r, j_r}$  at random according to the discrete distribution  $\{\alpha_{i,j}\}$

Select a state sample  $\bar{x}_{i_r, j_r}$  at random according to the predicted Gaussian distribution  $K_{i_r, j_r}$

End

The  $r$ th sample is associated with a specific prediction  $(i_r, j_r)$ , making it easy to  
5 apply the IEKF.

Given a set of starting points, we can apply the IEKF approach by modifying the IEKF equations to support linearization around arbitrary points. Consider a starting point  $\bar{x}$ . The nonlinear measurement model for template tracking can be written

$$z_t(x_t) = w_t$$

10 Expanding around  $\bar{x}$ , discarding high-order terms and rearranging gives

$$\bar{y}_t = z_t(\bar{x}) - \bar{C}_t \bar{x} = -\bar{C}_t x_t + w_t$$

where the auxiliary measurement  $\bar{y}_t$  has been defined to give a measurement model in standard form. It follows that the posterior update equation is

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + L_t(z_t(\bar{x}) + \bar{C}_t(\hat{x}_{t|t-1} - \bar{x})) \quad (\text{Eq. 29})$$

15 Note that if  $\bar{x} = x_{t|t-1}$ , which is the standard operating point for IEKF, then the above reduces to Equation 21. The additional term captures the effect of the new operating point. Equation 29 defines a modification of the IEKF update block to handle arbitrary starting points in computing the posterior update.

Note that the smoothness of the underlying nonlinear measurement model will  
20 determine the region in state space over which the linearized model  $\bar{C}_t$  is an accurate approximation. This region must be large enough relative to the distance  $\|\hat{x}_{t|t-1} - \bar{x}\|$ .

This requirement may not be met by a complex objective function.

An alternative in that case is to apply a standard gradient descent approach instead of IEKF, effectively discounting the role of the prior state prediction. A



multiple hypothesis tracking (MHT) approach which implements this procedure is described in Tat-Jen Cham and James M. Rehg, "A Multiple Hypothesis Approach to Figure Tracking," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 1999, Ft. Collins, CO., pages 239-245, incorporated herein by  
 5 reference in its entirety. Although this article does not describe tracking with an SLDS model or sampling from an SLDS prediction, the MHT procedure provides a means for propagating a set of sampled states given a complex measurement function. This means can be used as an alternative to the IEKF update step.

Fig. 18 is a block diagram of a tracking embodiment which combines SLDS  
 10 prediction with sampling from the prior mixture density to perform tracking. The output of the Viterbi predictor 510 follows two paths. The top path is similar to the embodiment of Fig. 16. The predictions are processed by a selector 515 and then input to an IEKF update block 506.

Along the bottom path in Fig. 18, the predictions are input to a sample generator  
 15 517, which produces a new set of R sample points for filtering,  $\{\bar{x}_{t,j_r}\}_{r=1}^R$ . This set is unioned with the SLDS predictions from the selector 515 and input to the IEKF block 506. The output of the IEKF block 506 is a joint set of filtered estimates, corresponding to the SLDS predictions, which we now write as  $\bar{x}_{t|i}$ , and the sampled states  $\bar{x}_{t|i,j_r}$ .

This combined output forms the input to a final selector 519 which selects one  
 20 filtered estimate for each of the switching states to make up the final output set. This selection process is identical to the ones described earlier with respect to Fig. 17, except that there now can be more than one possible estimate for a given state transition  $(i, j)$ , corresponding to different starting points for search.

Fig. 19 illustrates yet another tracking embodiment. The output of Viterbi  
 25 prediction, which comprises "Viterbi estimates," is input directly to an IEKF update block 506 while the output of the sample generator 517 goes to an MHT block 520, which implements the method of Cham and Rehg referred to above. As with the

007060-1045950

embodiment of Fig. 18, these two sets of filtered estimates are passed to a final selector 522. This final selector 522 chooses  $S$  of the posterior estimates, one for each possible switching state, as the final output. As with the embodiment of Fig. 18, this selector uses the posterior switching costs defined in Equation 25B.

5        It should be clear from the embodiments of Figs. 18 and 19 that other variations on the same basic approach are also possible. For example, an additional selector can be added following the predictor of Fig. 19, or the MHT block can be replaced by a second IEKF block.

10        It has been assumed that the selector would be selecting a single most likely state distribution from a set of distributions associated with a particular switching state. Another possibility is to find a single distribution which most closely matches all of the possibilities. This is the Generalized Psuedo Bayesian approximation GPB2, described earlier. In any of the selector blocks for the embodiments discussed above, GPB2 could be used in place of Viterbi approximation as a method for reducing multiple hypotheses  
15        for a particular switching state down to a single hypothesis. In cases where no single distribution contains a majority of the probability mass for the set of hypotheses, this approach may be advantageous. The application of the previously described process for GPB2 inference to these embodiments is straightforward.

#### Synthesis and Interpolation

20        SLDS was introduced as a “generative” model; trajectories in the state space can be easily generated by simulating an SLDS. Nevertheless, SLDSs are still more commonly employed as a classifier or a filter/predictor than as a generative model. We now formulate a framework for using SLDSs as synthesizers and interpolators.

25        Consider again the generative model described previously. Provided SLDS parameters have been learned from a corpus of motion trajectories, driving the generative SLDS model with the appropriate state and measurement noise processes and switching model, will yield a state space trajectory consistent with that corpus. In

other words, one will draw a sample (trajectory) defined by the probability distribution of the SLDS.

Fig. 20 illustrates a framework for synthesis of state space trajectories in which the SLDS is used as a generative model within a synthesis module 410. Given the parameters of the SLDS model 411, obtained using either SLDS learning or some other techniques, a switching state sequence  $s_t$  is first synthesized in the switching state synthesis module 412 by sampling from a Markov chain with state transition probability matrix  $\Pi$  and initial state distribution  $\pi_0$ . The continuous state sequence  $x_t$  is then synthesized in the continuous state synthesis module 413 by sampling from a LDS with parameters  $A(s_t)$ ,  $Q(s_t)$ ,  $C$ ,  $R$ ,  $x_0(s_0)$ , and  $Q_0(s_0)$ .

The above procedure will produce a random sequence of samples from the SLDS model. If an average noiseless state trajectory is desired, the synthesis can be run with LDS noise parameters (  $Q(s_t)$ ,  $R$ ,  $Q_0(s_0)$  ) set to zero and switching states whose duration is equal to average state durations, as determined by the switching state transition matrix  $\Pi$ . For example, this would result in sequences of prototypical walk or jog motions, whereas the random sampling would exhibit deviations from such prototypes. Intermediate levels of randomness can be achieved by scaling the SLDS model noise parameters with factors between 0 and 1.

The model parameters can also be modified to meet new constraints that were, for instance, not present in the data used to learn the SLDS. For example, initial state mean  $x_0$ , variance  $Q_0$  and switching state distribution  $\pi$  can be changed so as to force the SLDS to start in some arbitrary state  $x_a$  of regime  $i_a$ , e.g., to start simulation in a “walking” regime  $i_a$  with figure posture  $x_a$ . To achieve this, set  $x_0(i_a) = x_a$ ,  $Q_0(i_a) = 0$ ,  $\pi(i_a) = 1$ , and then proceed with the synthesis of this constrained model.

A framework of optimal control can be used to formalize synthesis under constraints. Optimal control of linear dynamic systems is described, for example, in B. Anderson and J. Moore, “Optimal Control: Linear Quadratic Methods,” Prentice Hall, Englewood Cliffs, NJ, 1990. For a LDS, the optimal control framework provides a way to design an optimal input or control  $u_t$  to the LDS, such that the LDS state  $x_t$  gets as close as possible to a desired state  $x_t^d$ . The desired states can also be viewed as

constraint points. The same formalism can be applied to SLDSs. Namely, the system equation, Equation 1, can be modified as

$$x_{t+1} = A(s_{t+1})x_t + u_{t+1} + v_{t+1}(s_{t+1}) \quad (\text{Eq. 1A})$$

Fig. 21 is a dependency graph of the modified SLDS 550 with added controls  $u_t$   
 5 552. A goal is to find  $u_t$  that makes  $x_t$  as close as possible to a given  $x_t^d$ . Usually, a quadratic measure of closeness is used, i.e., a control  $u_t$  is desired such that the cost  $V^{(x)}$ , or value function

$$V^{(x)} = \left\langle \sum_{t=0}^{T-1} \left[ (x_t - x_t^d)' W_t^{(x)} (x_t - x_t^d) + u_t' W_t^{(u)} u_t \right] \right\rangle$$

is minimized, where  $W_t^{(x)}$  and  $W_t^{(u)}$  are weight matrices. The optimal control is then

$$10 \quad \hat{u}_t = \arg \min_u V^{(x)}$$

For instance, if a SLDS is used to simulate a motion of the human figure,  $x_t^d$  might correspond to a desired figure posture at key frame  $t$ , and  $W_t^{(x)}$  might designate the key frame, i.e.,  $W_t^{(x)}$  is large for the key frame, and small otherwise.

In addition to “closeness” constraints, other types of constraints can similarly be  
 15 added. For example, one can consider a minimum-time constraint where the terminal timer  $T$  is to be minimized with the optimal control  $\hat{u}_t$ . In that case, the value function to be minimized becomes  $V^{(x)} \leftarrow V^{(x)} + T$ .

Other types of constraints that can be cast in this framework are the inequality or bounding constraints on the state  $x_t$  or control  $u_t$  (e.g.,  $x_t > x^{\min}$ ,  $u_t < u^{\max}$ ). Such  
 20 constraints could prevent, for example, the limbs of a simulated human figure from assuming physically unrealistic postures, or the control  $u_t$  from becoming unrealistically large.

Alternatively, as shown in Fig. 22, the SLDS 560 can be modified to include inputs 562, i.e., controls, to the switching state. The modified SLDS 560 is then described by the following equation:

$$\Pr(s_{t+1} = i | s_t = j, a_{t+1} = k) = \Pi^{(a)}(i, j, k)$$

Using the switching control  $u$ , constraints imposed on the switching states can be satisfied. Such constraints can be formulated similarly to constraints for the continuous control input of Fig. 21, e.g., switching state constraints, switching input constraints and minimum-time constraints. For example, a switching state constraint can guarantee that a figure is in the walking motion regime from time  $t_s$  to  $t_e$ . To find an optimal control  $\hat{a}_t$  that satisfies those constraints, one would have to use a modified value function that includes the cost of the switching state control. A framework for the switching state optimal control could be derived from the theory of reinforcement learning and Markov decision processes. See, for example, Sutton, R. S. and Barto, A. G., "Reinforcement Learning: An Introduction," Cambridge, MA, MIT Press, 1998.

$$V^{(s)} = - \left\langle \sum_{t=0}^{T-1} \gamma^t c_t(s_t, s_{t-1}, a_t) \right\rangle$$

Here,  $c_t(s_t, s_{t-1}, a_t)$  represents a cost for making the transition from switching state  $s_{t-1}$  to  $s_t$ , for a given control  $a_t$ , and  $\gamma$  is a discount or “forgetting” factor. The cost

function  $c_i$  is designed to emphasize, with a low  $c_i$ , those states that agree with imposed constraints, and to penalize, with a high  $c_i$ , those states that violate the imposed constraints. Once the optimal control  $\hat{a}_i$  has been designed, the modified generative SLDS model can be driven by noise to produce a synthetic trajectory.

- 5 As Fig. 23 illustrates, the SLDS system 750 can be modified to include both types of controls, continuous 572 and switching 574, as indicated by the following equation.

$$x_{t+1} = A(s_{t+1})x_t + u_{t+1} + v(s_{t+1})$$

$$\Pr(s_{t+1} = i | s_t = j, a_{t+1} = k) = \Pi^{(a)}(i, j, k)$$

- 10 In this model 570, the mixed control  $(u, a)$  can lead to both a desired switching state, e.g., motion regime, and a desired continuous state, e.g., figure posture. For example, a constraint can be specified that requires the human figure to be in the walking regime  $i_a$  with some specific posture  $x_a$  at time  $t$ . As in the case of the continuous and switching state optimal controls, additional constraints such as input bounding and minimum time can also be specified for the mixed state control. To find
- 15 the optimal control  $(\hat{u}_t, \hat{a}_t)$ , a value function can be used that includes the costs of the switching and the continuous state controls, e.g.,  $V^{(x)} + V^{(s)}$ . Again, once the optimal control  $(\hat{u}_t, \hat{a}_t)$  is designed, the modified generative SLDS model can be used to produce a synthetic trajectory.

- 20 Fig. 24 illustrates the framework 580, for synthesis under constraints, which utilizes optimal control. The SLDS model 582 is modified by a SLDS model modification module 584 to include the control terms or inputs. Using the modified model 585, an optimal control module 586 finds the optimal controls 587 which satisfy

constraints 588. Finally, a synthesis model 590 generates synthesized trajectories 592 from the modified SLDS 585 and the optimal controls 587.

5 The use of optimal controls by the present invention to generate motion by sampling from SLDS models in the presence of constraints has broad applications in computer animation and computer graphics. The task of generating realistic or compelling motions for synthetic characters has long been recognized as an extremely challenging problem. One classical approach to this problem is based on the notion of spacetime constraints which was first introduced by Andrew Witkin and Michael Kass, "Spacetime Constraints," Computer Graphics, Volume 22, Number 4, August 1988

10 pages 159-168. In this approach, an optimal control problem is formulated over an analytic dynamical model derived from Newtonian physics. For example, in order to make a lamp hop in a realistic way, the animator would derive the physical equations which govern the lamp's motion. These equations involve the mass distribution of the lamp and forces such as gravity that are acting upon it.

15 Unfortunately, this method of animation has proved to be extremely difficult to use in practice. There are two main problems. First, it is extremely difficult to specify all of the equations and parameters that are necessary to produce a desired motion. In the case of the human figure, for example, specifying all of the model parameters required for realistic motion is a daunting task. The second problem is that the resulting

20 equations of motion are highly complex and nonlinear, and usually involve an enormous number of state variables. For example, the jumping Luxo lamp described in the Witkin and Kass paper involved 223 constraints and 394 state variables. The numerical methods which must be used to solve control problems in such a large complex state space are also difficult to work with. Their implementation is complex

25 and their convergence to a correct solution can be problematic.

In contrast, our approach of optimal control of learned SLDS models has the potential to overcome both of these drawbacks. By using a learned switching model, desired attributes such as realism can be obtained directly from motion data. Thus, there is no need to specify the physics of human motion explicitly in order to obtain

5

10

20

In contrast, sampling repeatedly from our SLDS framework can produce motions which differ in their small details but are qualitatively consistent. This type of randomness is necessary in order to avoid awkward repetitions in an animation application. Furthermore, the learning approach makes it possible to generalize from a



set of motions to new motions that have not been seen before. In contrast, the approach of retargeting is based on modifying a single instance of motion data.

Another approach to synthesizing animations from learned models is described in Matthew Brand, "Voice Puppetry," SIGGRAPH 99 Conference Proceedings, Annual  
5 Conference Series 1999, pages 21-28 and in Matthew Brand and Aaron Hertzmann, "Style Machines," SIGGRAPH 2000 Conference Proceedings, Annual Conference Series 2000, pages 183-192. This method uses sampling from a Hidden Markov Model to synthesize facial and figure animations learned from training data. Unlike our SLDS  
10 framework, the HMM representation is limited to using piecewise constant functions to model the feature data during learning. This can require a large number of discrete states in order to capture subtle effects when working with complex state space data.

In contrast, our framework employs a set of LDS models to describe feature data, resulting in models with much more expressive power. Furthermore, the prior art  
15 does not describe any mechanism for imposing constraints on the samples from the models. This may make it difficult to use this approach in achieving specific animation objectives.

To test the power of the learned SLDS synthesis/interpolation framework, we examined its use in synthesizing realistic-looking motion sequences and interpolating motion between missing frames. In one set of experiments, the learned walk/jog SLDS  
20 was used to generate a "synthetic walk" based on initial conditions learned by the SLDS model.

Fig. 25 illustrates a stick figure 220 motion sequence of the noise driven model. Depending on the amount of noise used to drive the model, the stick figure exhibits more or less "natural"-looking walk. Departure from the realistic walk becomes more  
25 evident as the simulation time progresses. This behavior is not unexpected as the SLDS in fact learns locally consistent motion patterns. Fig. 25 illustrates a synthesized walk motion over 50 frames using SLDS as a generative model. The states of the synthesized motion are shown on the graph 222.

0918:1305-000

Another realistic situation may call for filling in a small number of missing frames from a large motion sequence. SLDS can then be utilized as an interpolation function. In another set of experiments, we employed the learned walk/jog model to interpolate a walk motion over two sequences with missing frames. Missing-frame  
5 constraints were included in the interpolation framework by setting the measurement variances corresponding to those frames to infinity. The visual quality of the interpolation and the motion synthesized from it was high. As expected, the sparseness of the measurement set had definite bearing on this quality.

#### USE OF THE INVENTION

10 Our invention makes possible a number of core tasks related to the analysis and synthesis of the human figure motion:

- Track figure motion in an image sequence using learned dynamic models.
- Classify different types of human motion.
- Synthesize motion using stochastic models that correspond to different types of  
15 motion.
- Interpolate missing motion data from sparsely observed image sequences.

We anticipate that our invention could impact the following application areas:

- **Surveillance:** Use of accurate dynamic models could lead to improved tracking in noisy video footage. The ability to interpolate missing data could be useful in  
20 situations where frame rates are low, as in Web or other network applications. The ability to classify motion into categories can benefit from the SLDS approach. Two forms of classification are possible. First, specific actions such as “opening a door” or “dropping a package” can be modeled using our approach. Second, it may be possible to recognize specific individuals based on  
25 the observed dynamics of their motion in image sequences. This could be used, for example, to recognize criminal suspects using surveillance cameras in public places such as airports or train stations.

- **User-interfaces:** Interfaces based on vision sensing could benefit from improved tracking and better classification performance due to the SLDS approach.
- **Motion capture:** Motion capture in unstructured environments can be enabled through better tracking techniques. In addition to the capture of live motion without the use of special clothing, it is also possible to capture motion from archival sources such as old movies.
- **Motion synthesis:** The generation of human motion for computer graphics animation can be enabled through the use of a learned, generative stochastic model. By learning models from sample motions, it is possible to capture the natural dynamics implicit in real human motion without a laborious manual modeling process. Because the resulting models are stochastic, sampling from the models produces motion with a pleasing degree of randomness.
- **Video editing:** Tracking algorithms based on powerful dynamic models can simplify the task of segmenting video sequences.
- **Video compression/decompression:** The ability to interpolate a video sequence based on a sparse set of samples could provide a new approach to coding and decoding video sequences containing human or other motion. In practice, human motion is common in video sequences. By transmitting key frames detected using SLSDS classification at a low sampling rate and interpolating, using SLDS interpolation, the missing frames from the transmitted model parameters, a substantial savings in bit-rate may be achievable.

It will be apparent to those of ordinary skill in the art that methods involved in the present system for a method for motion synthesis and interpolation using switching linear dynamic system models may be embodied in a computer program product that includes a computer usable medium. For example, such a computer usable medium can include a readable memory device, such as a hard drive device, a CD-ROM, a DVD-

ROM, or a computer diskette, having computer readable program code segments stored thereon. The computer readable medium can also include a communications or transmission medium, such as a bus or a communications link, either optical, wired, or wireless, having program code segments carried thereon as digital or analog data  
5 signals.

While this invention has been particularly shown and described with references to preferred embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the scope of the invention encompassed by the appended claims.

0918.1305-000

## CLAIMS

What is claimed is:

1. A method for synthesizing a sequence, comprising:
  - defining a switching linear dynamic system (SLDS) comprising a plurality of dynamic models;
  - associating each model with a switching state such that a model is selected when its associated switching state is true;
  - determining a state transition record for at least one training sequence of measurements by determining and recording, for a given measurement and for each possible switching state, an optimal prior switching state, based on the at least one training sequence, wherein the optimal prior switching state optimizes a transition probability;
  - determining, for a final measurement, an optimal final switching state;
  - determining the sequence of switching states by backtracking, from said optimal final switching state, through the state transition record;
  - learning parameters of the dynamic models, responsive to the determined sequence of switching states; and
  - synthesizing a new data sequence, based on the dynamic models with learned parameters.
2. The method of Claim 1, wherein the new data sequence has characteristics which are similar to characteristics of at least one training sequence.
3. The method of Claim 1, wherein the new data sequence combines characteristics of plural training sequences.
4. The method of Claim 1, further comprising modifying the SLDS such that at least one constraint is met.



14. The method of Claim 12, further comprising:  
designing an optimal switching control that satisfies constraints.
15. The method of Claim 14, further comprising:  
synthesizing the new data sequence using the optimal control.
- 5 16. The method of Claim 4, further comprising designing optimal switching and  
continuous state controls that satisfy continuous and switching constraints  
respectively.
17. The method of Claim 16, further comprising:  
synthesizing the new data sequence using the optimal controls.
- 10 18. The method of Claim 1, wherein the sequence of measurements comprises  
economic data.
19. The method of Claim 1, wherein the sequence of measurements comprises  
image data.
20. The method of Claim 1, wherein the sequence of measurements comprises audio  
15 data.
21. The method of Claim 1, wherein the sequence of measurements comprises  
spatial data.
22. A switching linear dynamic system (SLDS) model, comprising:  
a plurality of linear dynamic system (LDS) models, wherein at any given  
20 instance, an LDS model is selected responsive to a switching variable;  
a state transition recorder which determines, for at least one training

5

10

10

15

- 15

- 20

- 25

- means for associating each model with a switching state such that a model is selected when its associated switching state is true;

- means for determining a state transition record for at least one training sequence of measurements by determining and recording, for a given measurement and for each possible switching state, an optimal prior switching state, based on the at least one sequence, wherein the optimal prior switching state optimizes a transition probability;



means for determining, for a final measurement, an optimal final switching state;

means for determining the sequence of switching states by backtracking, from said optimal final switching state, through the state transition record;

5 means for learning parameters of the dynamic models, responsive to the determined sequence of switching states; and

means for synthesizing a new data sequence, based on the dynamic models with learned parameters.

26. A computer program product for synthesizing a sequence, the computer  
10 program product comprising a computer usable medium having computer readable code thereon, including program code which:
- defines a switching linear dynamic system (SLDS) comprising a plurality of dynamic models;
  - associates each model with a switching state such that a model is  
15 selected when its associated switching state is true;
  - determines a state transition record for at least one training sequence of measurements by determining and recording, for a given measurement and for each possible switching state, an optimal prior switching state, based on the at least one training sequence, wherein the optimal prior switching state optimizes  
20 a transition probability;
  - determines an optimal final switching state for a final measurement; and
  - determines the sequence of switching states by backtracking, from said optimal final switching state, through the state transition record;
  - learns parameters of the dynamic models, responsive to the determined  
25 sequence of switching states; and
  - synthesizes a new data sequence, based on the dynamic models with learned parameters.

27. A computer system comprising:
- a processor;
  - a memory system connected to the processor; and
  - a computer program, in the memory, which:
    - 5 associates each of a plurality of dynamic models with a switching state such that a model is selected when its associated switching state is true;
    - determines a state transition record for at least one training sequence of measurements by determining and recording, for a given measurement and for each possible switching state, an optimal prior switching state, based on the at
    - 10 least one training sequence, wherein the optimal prior switching state optimizes a transition probability;
    - determines an optimal final switching state for a final measurement; and
    - determines the sequence of switching states by backtracking, from said optimal final switching state, through the state transition record;
    - 15 learns parameters of the dynamic models, responsive to the determined sequence of switching states; and
    - synthesizes a new data sequence, based on the dynamic models with learned parameters.
28. A computer data signal embodied in a carrier wave for synthesizing a sequence,
- 20 comprising:
- program code for associating each model with a switching state such that a model is selected when its associated switching state is true;
  - program code for determining a state transition record for at least one training sequence of measurements by determining and recording, for a given
  - 25 measurement and for each possible switching state, an optimal prior switching state, based on the at least one training sequence, wherein the optimal prior switching state optimizes a transition probability;
  - program code for determining, for a final measurement, an optimal final

switching state;

program code for determining the sequence of switching states by backtracking, from said optimal final switching state, through the state transition record; and

5           program code for learning parameters of the dynamic models, responsive  
to the determined sequence of switching states; and

program code for synthesizing a new data sequence, based on the dynamic models with learned parameters.

29. A method for synthesizing a sequence, comprising:

10 defining a switching linear dynamic system (SLDS) comprising a plurality of dynamic models;

associating each dynamic model with a switching state such that a dynamic model is selected when its associated switching state is true, wherein the switching state at a particular instance is determined by a switching model;

15           decoupling the dynamic models from the switching model;

determining parameters of a decoupled dynamic model, responsive to a switching state probability estimate;

estimating a state of a decoupled dynamic model corresponding to a measurement at the particular instance, and responsive to at least one training sequence of measurements;

determining parameters of the decoupled switching model, responsive to the dynamic state estimate;

estimating a probability for each possible switching state of the decoupled switching model;

25                   determining a switching state sequence based on the estimated switching  
state probabilities;

learning parameters of the dynamic models, responsive to the switching states sequence; and

synthesizing a new data sequence, based on the dynamic models with learned parameters.

30. The method of Claim 29, wherein the new data sequence has characteristics similar to at least one training sequence.
- 5 31. The method of Claim 29, wherein the new data sequence combines characteristics of plural training sequences.
32. The method of Claim 32, further comprising modifying the SLDS such that at least one constraint is met.
- 10 33. The method of Claim 32, wherein modifying the SLDS comprises:  
adding a continuous state control.
34. The method of Claim 33, wherein modifying the SLDS further comprises:  
adding constraints on continuous states.
35. The method of Claim 33, wherein modifying the SLDS further comprises:  
adding constraints on the continuous state control.
- 15 36. The method of Claim 33, wherein modifying the SLDS further comprises:  
adding constraints on time.
37. The method of Claim 33, further comprising:  
designing an optimal continuous control that satisfies the at least one constraint.



an approximate variational state sequence inference module, which reestimates parameters of each LDS model, using variational inference, to minimize a modeling cost of current state sequence estimates, responsive to at least one training sequence of measurements; and

5 a synthesizer which synthesizes a new data sequence, based on the  
reestimated dynamic models.

47. The SLDS model of Claim 46, wherein the new data sequence has characteristics similar to the at least one training sequence.

48. The SLDS model of Claim 46, wherein the new data sequence combines  
10 characteristics of plural training sequences.

49. The method of Claim 46, further comprising modifying the SLDS such that at least one constraint is met.

50. The method of Claim 49, wherein modifying the SLDS comprises:  
adding a continuous state control.

15    51.    The method of Claim 49, wherein modifying the SLDS comprises:  
              adding a switching state control.

52. The method of Claim 49, further comprising designing optimal switching and continuous state controls that satisfy continuous and switching constraints respectively.

20 53. The method of Claim 52, further comprising:  
synthesizing the new data sequence using the optimal controls.

54. A method for interpolating from an input measurement sequence, comprising:  
defining a switching linear dynamic system (SLDS) comprising a plurality of dynamic models;  
associating each model with a switching state such that a model is  
5 selected when its associated switching state is true;  
determining a state transition record by determining and recording, for a given measurement and for each possible switching state, an optimal prior switching state, based on at least one training measurement sequence, wherein the optimal prior switching state optimizes a transition probability;  
10 determining, for a final measurement, an optimal final switching state;  
determining the sequence of switching states by backtracking, from said optimal final switching state, through the state transition record;  
determining the sequence of continuous states based on the determined sequence of switching states; and  
15 interpolating missing motion data from the input sequence, based on dynamic models and responsive to the determined sequences of continuous and switching states.
55. The method of Claim 54, further comprising modifying the SLDS such that at least one constraint is met.
- 20 56. The method of Claim 55, wherein modifying the SLDS comprises:  
adding a continuous state control.
57. The method of Claim 55, wherein modifying the SLDS comprises:  
adding a switching state control.

0918'1305-000

58. The method of Claim 55, further comprising designing optimal switching and continuous state controls that satisfy continuous and switching constraints respectively.
59. The method of Claim 58, further comprising:  
5 interpolating the new data sequence using the optimal controls.
60. The method of Claim 54, further comprising:  
at a receiver, interpolating missing frames from transmitted model parameters and from received key frames, the key frames having been determined based on the learned parameters, wherein the input measurement  
10 sequence comprises the received key frames.
61. A switching linear dynamic system (SLDS) model, comprising:  
a plurality of linear dynamic system (LDS) models, wherein at any given instance, an LDS model is selected responsive to a switching variable;  
a state transition recorder which determines a state transition record for a  
15 training measurement sequence by determining and recording, for a given measurement and for each possible switching state, an optimal prior switching state, based on the training sequence, wherein the optimal prior switching state optimizes a transition probability, and which determines, for a final measurement, an optimal final switching state;  
20 a backtracer which determines a sequence of switching states corresponding to the training sequence by backtracking, from said optimal final switching state, through the state transition record;  
a dynamic model learner which learns parameters of the dynamic models responsive to the determined sequence of switching states; and  
25 an interpolator which interpolates missing motion data from the input sequence, based on the dynamic models with learned parameters.



62. A system for interpolating from an input measurement sequence, the system comprising:
- means for defining a switching linear dynamic system (SLDS) comprising a plurality of dynamic models;
  - 5 means for associating each model with a switching state such that a model is selected when its associated switching state is true;
  - means for determining a state transition record by determining and recording, for a given measurement and for each possible switching state, an optimal prior switching state, based on at least one training sequence, wherein
  - 10 the optimal prior switching state optimizes a transition probability;
  - means for determining, for a final measurement, an optimal final switching state;
  - means for determining the sequence of switching states by backtracking, from said optimal final switching state, through the state transition record;
  - 15 means for learning parameters of the dynamic models, responsive to the determined sequence of switching states; and
  - means for interpolating missing motion data from the input sequence, based on dynamic models learned from training sequences.
63. A computer program product for interpolating from an input measurement
- 20 sequence, the computer program product comprising a computer usable medium having computer readable code thereon, including program code which:
- associates each model with a switching state such that a model is selected when its associated switching state is true;
  - determines a state transition record by determining and recording, for a
  - 25 given measurement and for each possible switching state, an optimal prior switching state, based on at least one training measurement sequence, wherein the optimal prior switching state optimizes a transition probability;
  - determines an optimal final switching state for a final measurement; and

0918.1305-000

determines the sequence of switching states by backtracking, from said optimal final switching state, through the state transition record;

learns parameters of the dynamic models, responsive to the determined sequence of switching states resulting; and

5 interpolates missing motion data from the input sequence, based on the dynamic models with learned parameters.

64. A computer system comprising:

a processor;

a memory system connected to the processor; and

10 a computer program, in the memory, which:

associates each of a plurality of dynamic models with a switching state such that a model is selected when its associated switching state is true;

determines, from a set of possible switching states and responsive to a training sequence of measurements, a state transition record by determining and recording, for a given measurement and for each possible switching state, an optimal prior switching state, wherein the optimal prior switching state optimizes a transition probability;

determines an optimal final switching state for a final measurement;

20 determines a sequence of switching states corresponding to the measurement sequence by backtracking, from said optimal final switching state, through the state transition record;

learns parameters of the dynamic models, responsive to the determined sequence of switching states; and

25 interpolates missing motion data from an input sequence, based on the dynamic models with learned parameters.

65. A computer data signal embodied in a carrier wave for interpolating from an input measurement sequence, comprising:

0918.1305-000

program code for associating each model with a switching state such that a model is selected when its associated switching state is true;

program code for determining a state transition record by determining and recording, for a given measurement of at least one training sequence and for each possible switching state, an optimal prior switching state, based on the at least one training sequence, wherein the optimal prior switching state optimizes a transition probability;

program code for determining, for a final measurement, an optimal final switching state;

program code for determining the sequence of switching states by backtracking, from said optimal final switching state, through the state transition record; and

program code for learning parameters of the dynamic models, responsive to the determined sequence of switching states; and

program code for interpolating missing data from an input sequence, based on the dynamic models with learned parameters.

66. A method for interpolating from an input measurement sequence, comprising:

defining a switching linear dynamic system (SLDS) comprising a plurality of dynamic models;

associating each dynamic model with a switching state such that a dynamic model is selected when its associated switching state is true, wherein the switching state at a particular instance is determined by a switching model;

decoupling the dynamic models from the switching model;

determining parameters of a decoupled dynamic model, responsive to a switching state probability estimate;

estimating a state of a decoupled dynamic model corresponding to a measurement at the particular instance, and responsive to at least one training sequence of measurements;

0054401-09100

determining parameters of the decoupled switching model, responsive to the dynamic state estimate;

estimating a probability for each possible switching state of the decoupled switching model; and

5 determining the sequence of switching states based on the estimated switching state probabilities;

learning parameters of the dynamic models, responsive to the determined sequence of switching states; and

10 interpolating missing motion data from the input sequence, based on the dynamic models with learned parameters.

67. The method of Claim 66, further comprising modifying the SLDS such that at least one constraint is met.

68. The method of Claim 67, wherein modifying the SLDS comprises:  
adding a continuous state control.

15 69. The method of Claim 67, wherein modifying the SLDS comprises:  
adding a switching state control.

70. The method of Claim 67, further comprising designing optimal switching and continuous state controls that satisfy continuous and switching constraints respectively.

20 71. The method of Claim 70, further comprising:  
synthesizing the new data sequence using the optimal controls.

72. The method of Claim 66, wherein the measurement sequence comprises a sparsely observed image sequence.

0918.1305-000

73. The method of Claim 66, further comprising:
- at a receiver, interpolating missing frames from transmitted model parameters and from received key frames, the key frames having been determined based on the learned parameters.
- 5 74. A switching linear dynamic system (SLDS) model, comprising:
- a plurality of linear dynamic system (LDS) models, wherein at any given instance, an LDS model is selected responsive to a switching variable;
- a switching model which determines values of the switching variable;
- an approximate variational state sequence inference module, which
- 10 reestimates parameters of each SLDS model, using variational inference, to minimize a modeling cost of current state sequence estimates;
- a dynamic model learner which learns parameters of the dynamic models responsive to the determined sequence of switching states resulting from at least one training sequence; and
- 15 an interpolator which interpolates missing motion data from an input sequence, based on the dynamic models with learned parameters.

METHOD FOR MOTION SYNTHESIS AND INTERPOLATION  
USING SWITCHING LINEAR DYNAMIC SYSTEM MODELS

ABSTRACT OF THE DISCLOSURE

5 A method for synthesizing a sequence includes defining a switching linear  
dynamic system (SLDS) with a plurality of dynamic systems. In a Viterbi-based  
method, a state transition record for a training sequence is determined. The  
corresponding sequence of switching states is determined by backtracking through the  
state transition record. Parameters of the dynamic models are learned in response to the  
determined sequence of switching states, and a new data sequence is synthesized, based  
10 on the dynamic models whose parameters have been learned. In a variational-based  
method, the switching state at a particular instance is determined by a switching model.  
The dynamic models are decoupled from the switching model, and parameters of the  
decoupled dynamic model are determined responsive to a switching state probability  
estimate. A state of a decoupled dynamic model corresponding to a measurement at the  
15 particular instance is estimated, responsive to one or more training sequences.  
Parameters of the decoupled switching model are then determined, responsive to the  
dynamic state estimate. A probability is estimated for each possible switching state of  
the decoupled switching model. The sequence of switching states is determined based  
on the estimated switching state probabilities. Parameters of the dynamic models are  
20 learned responsive to the determined sequence of switching states, and a new data  
sequence is synthesized based on the dynamic models with learned parameters. Similar  
methods are used to interpolate from an input sequence.

0918:1305-000



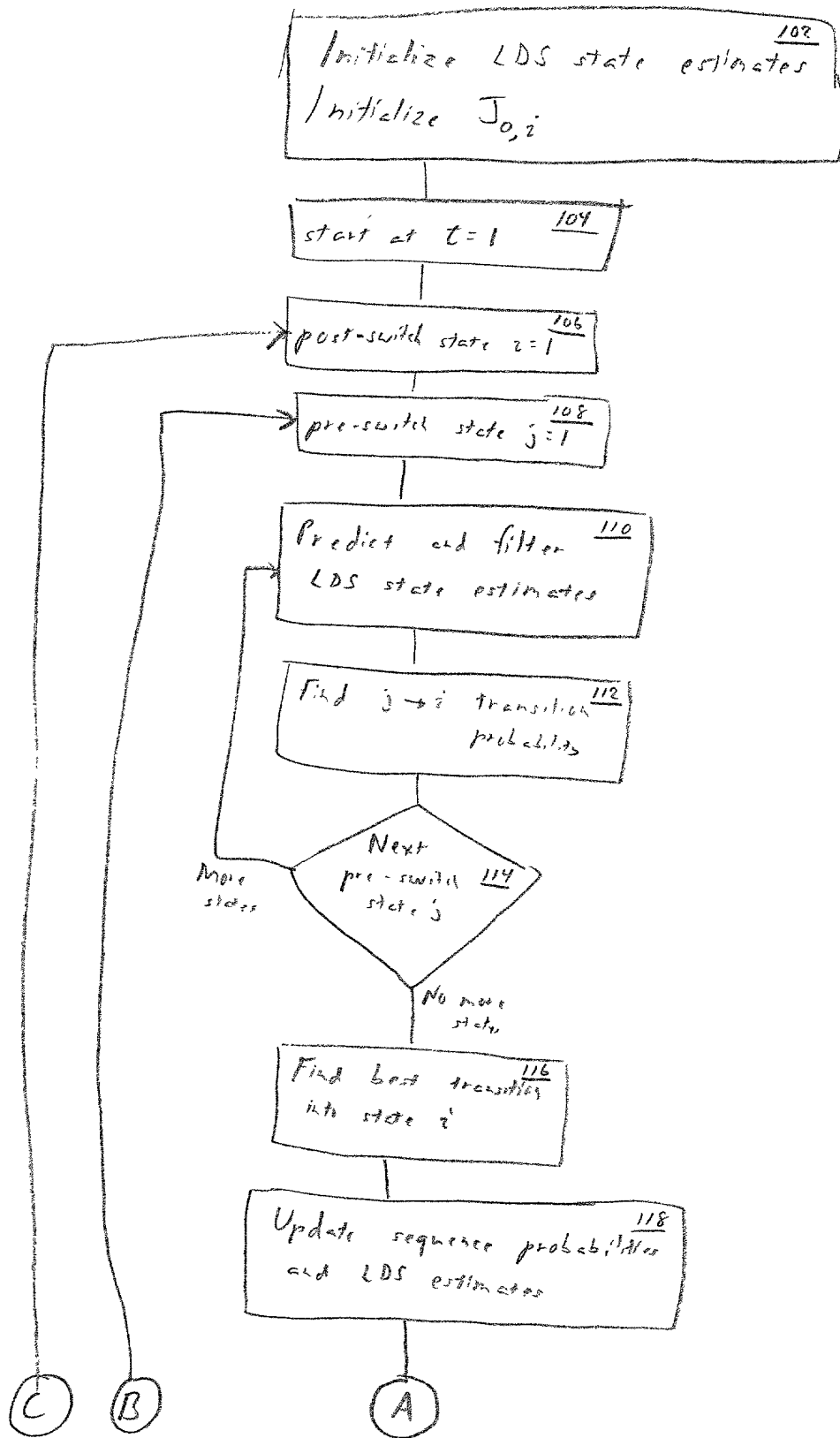


Fig. 4A



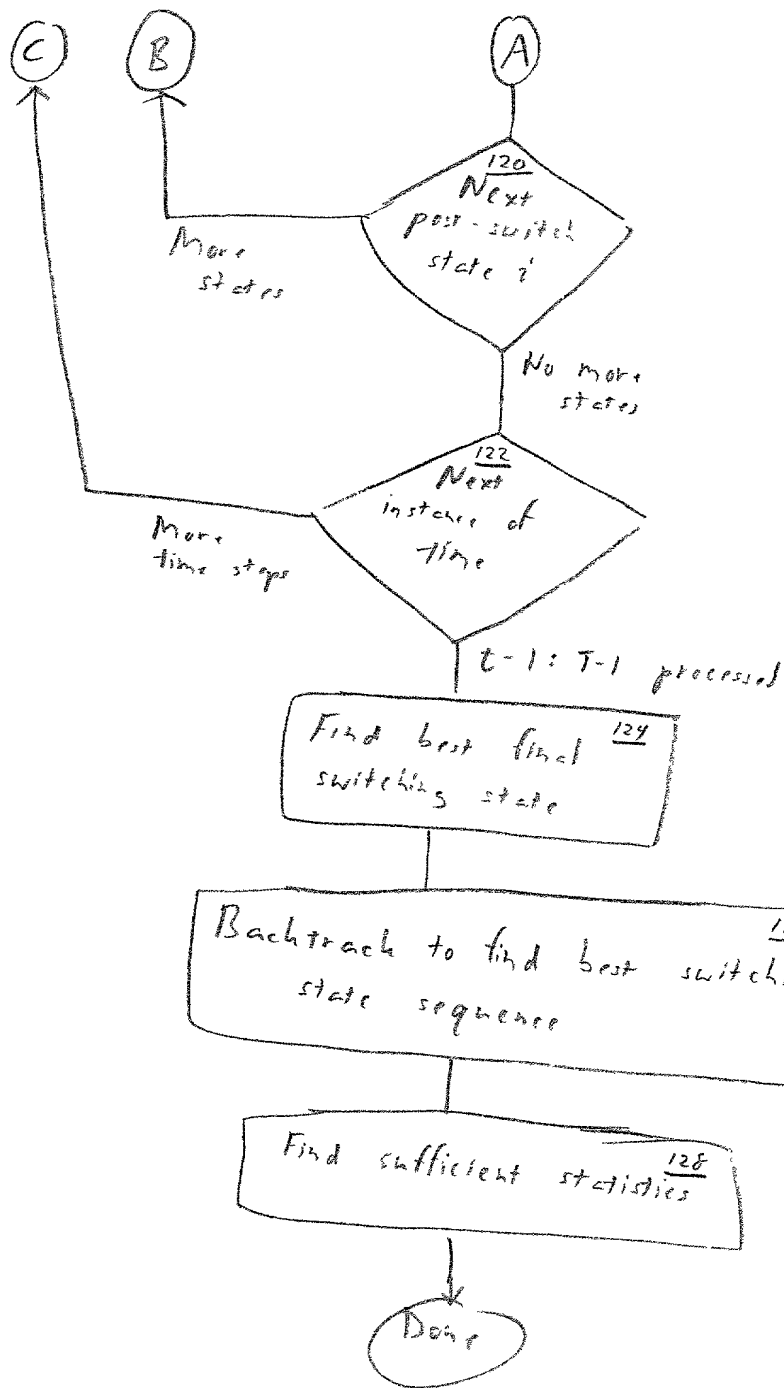


Fig. 4B

09054401.090103



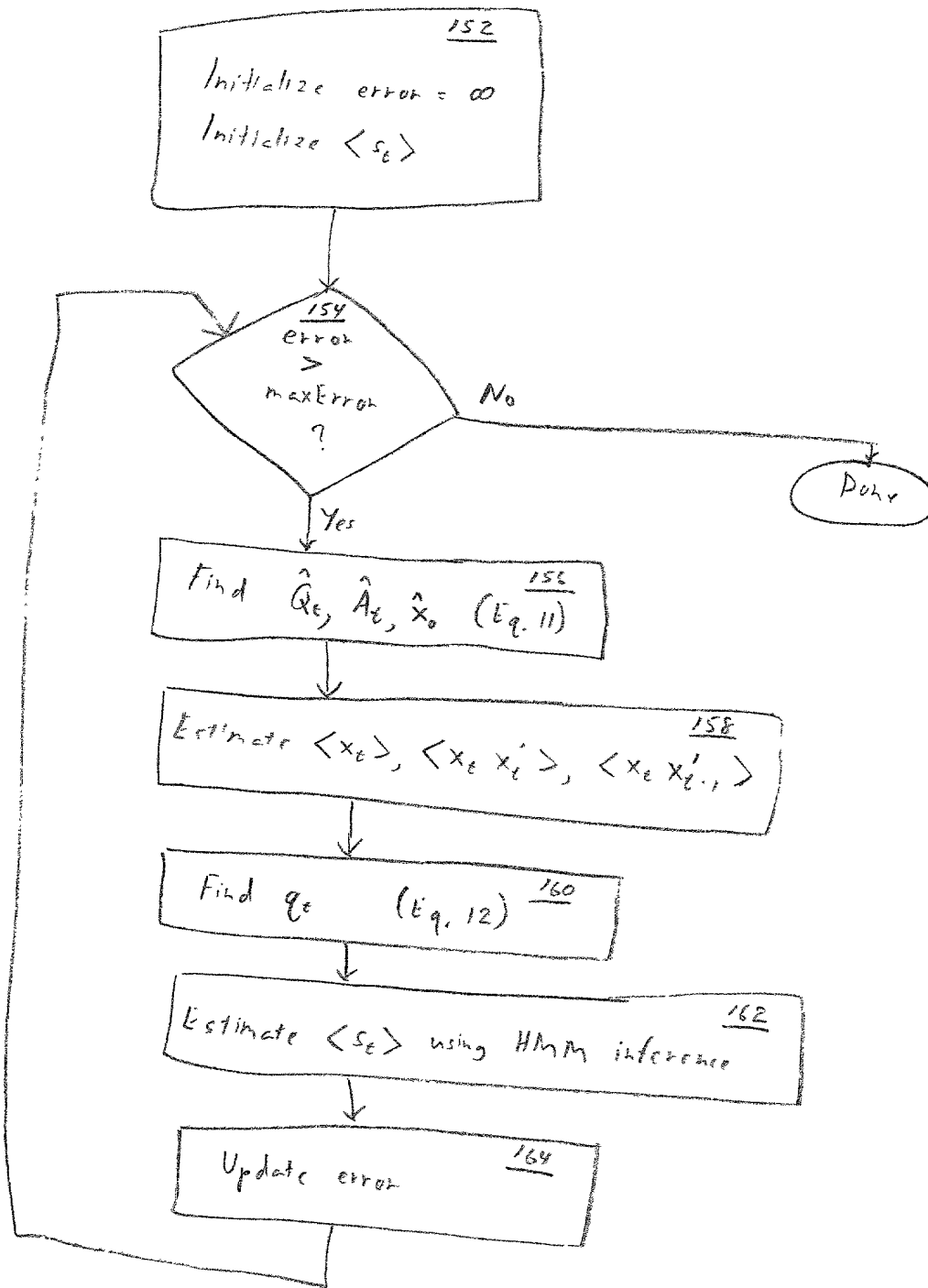


Fig. 7





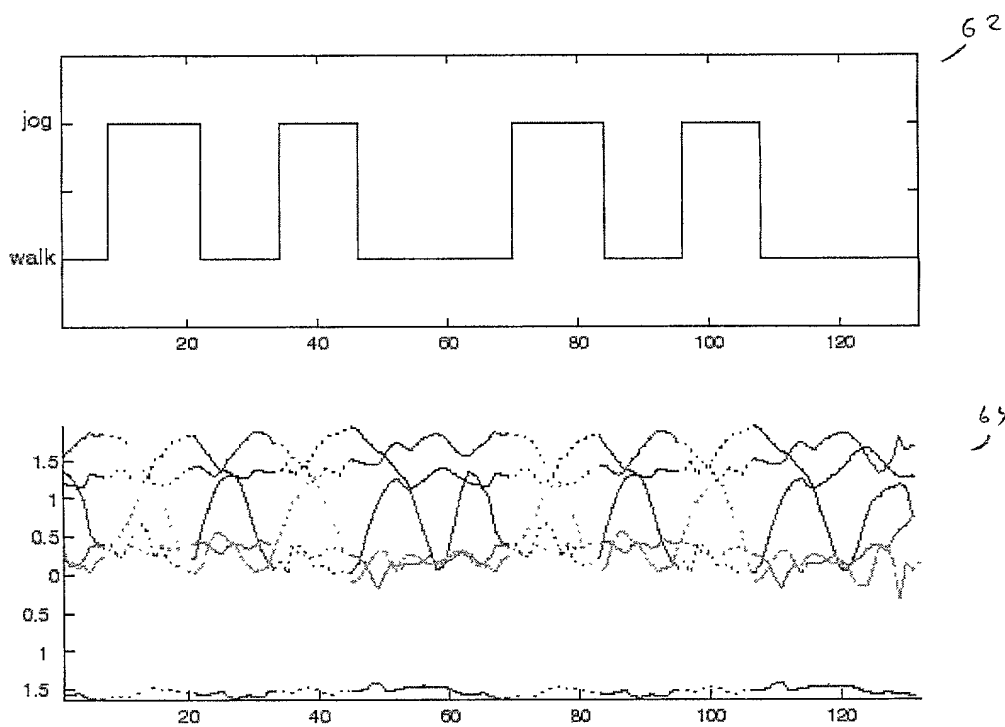


Fig. 9

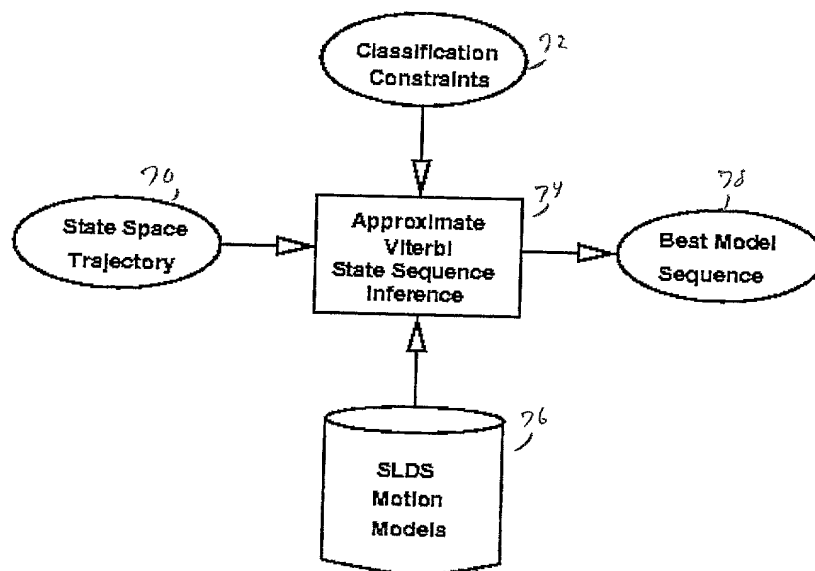


Fig. 10

005440.090100  
00T060" T045960

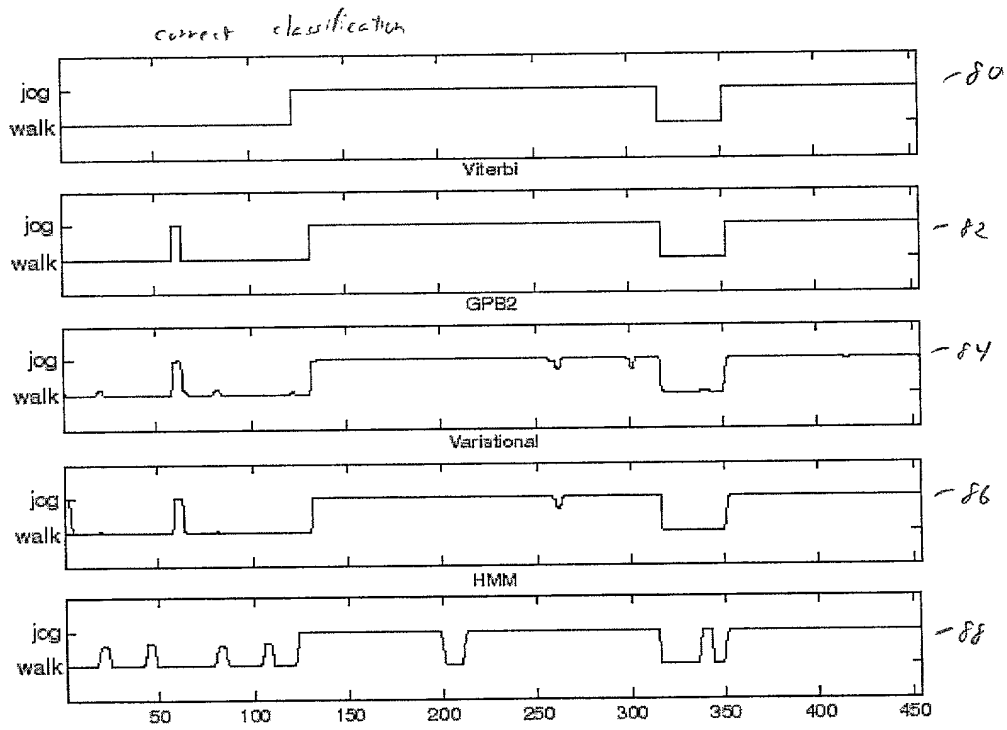


Fig 11





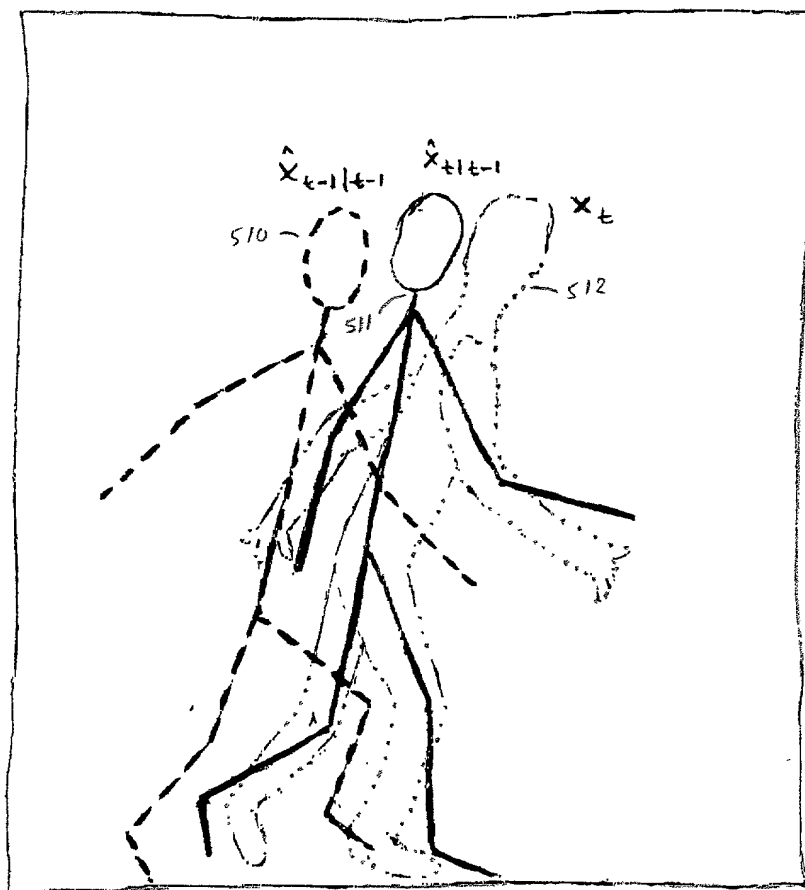


Fig. 13

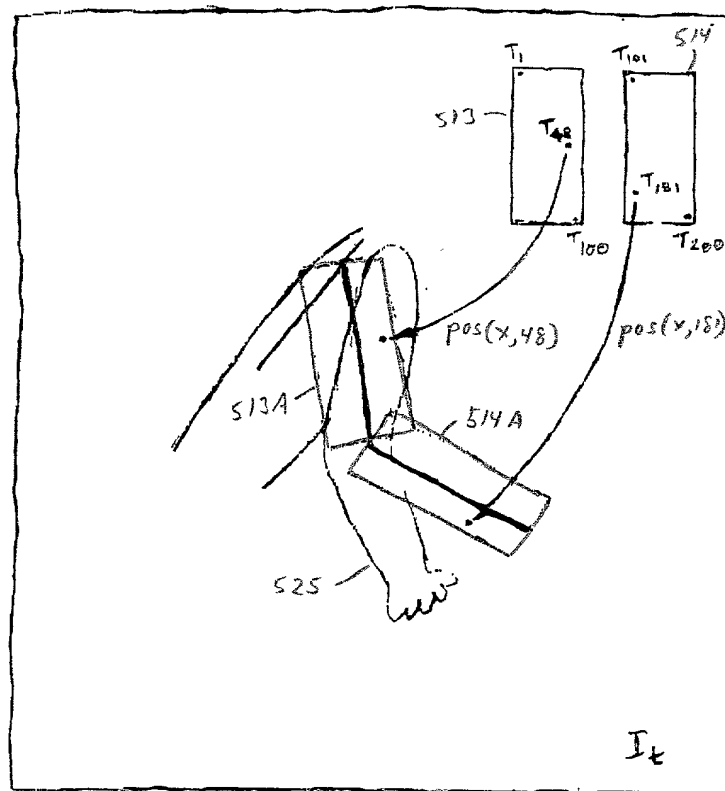


Fig. 14

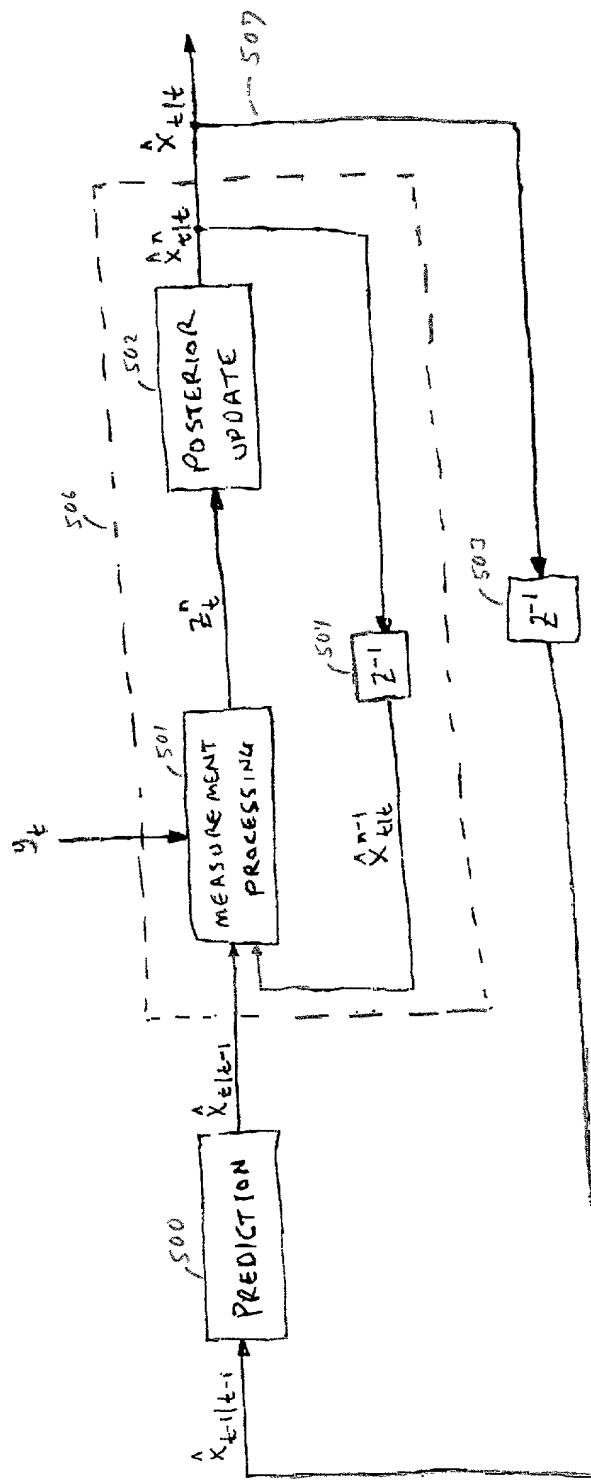


Fig. 15  
(Prior ART)

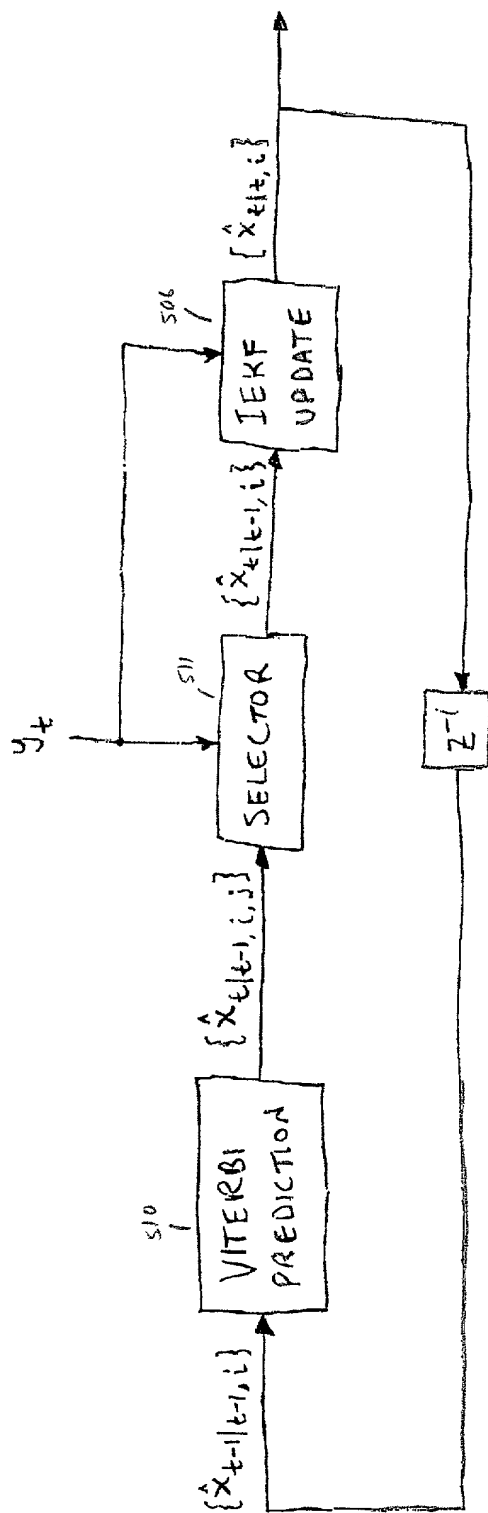


Fig. 16

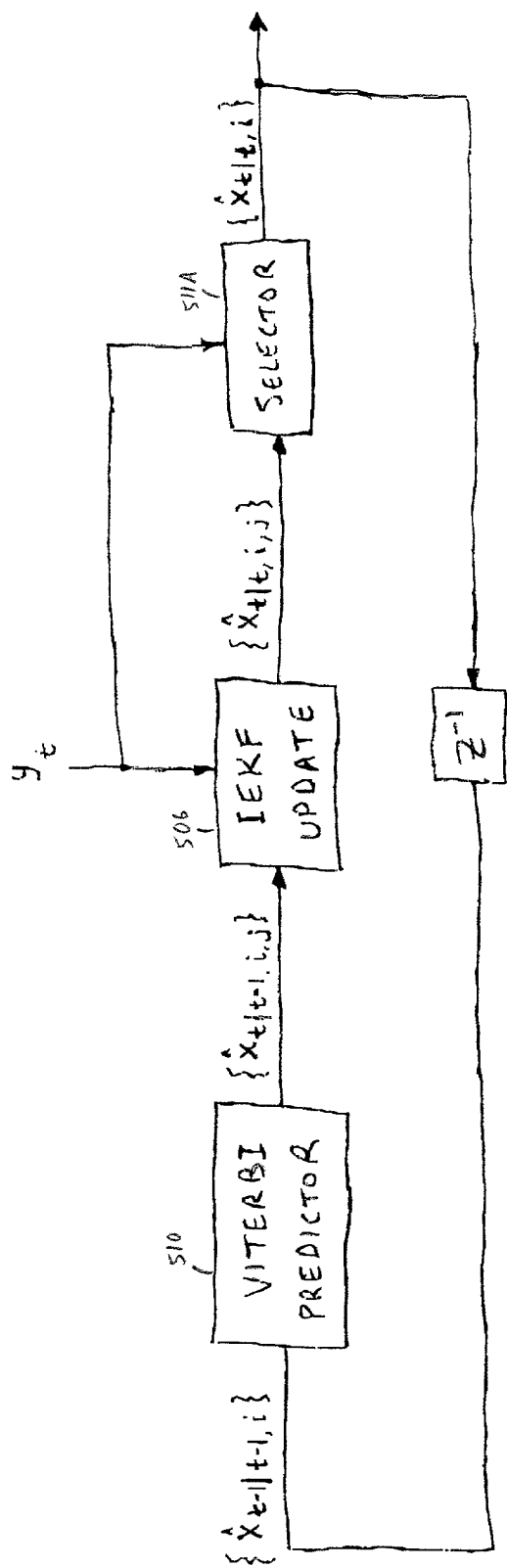


Fig. 17

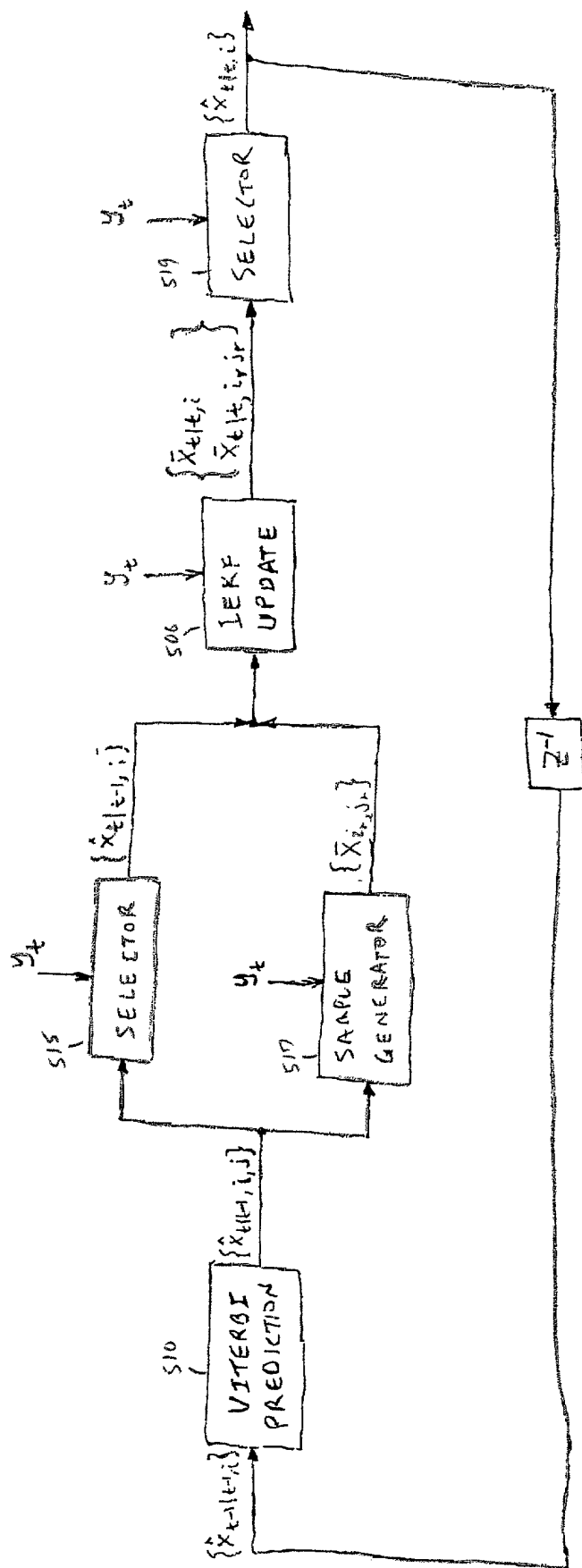


Fig. 18

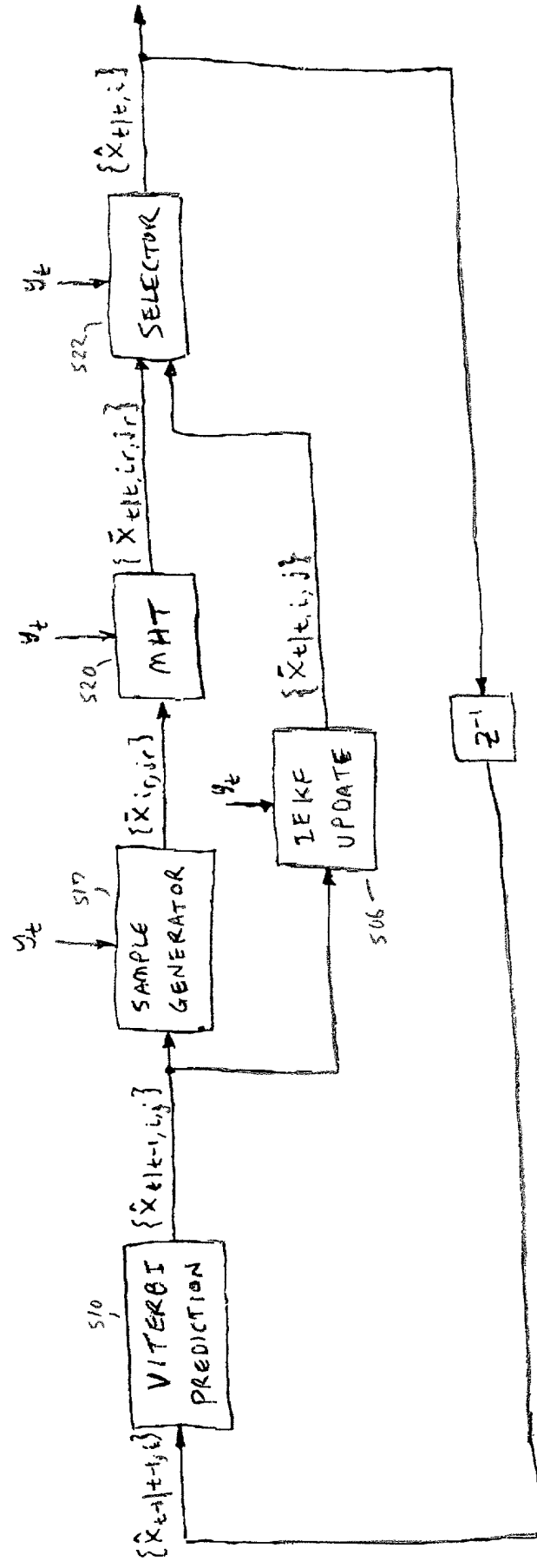


Fig. 19

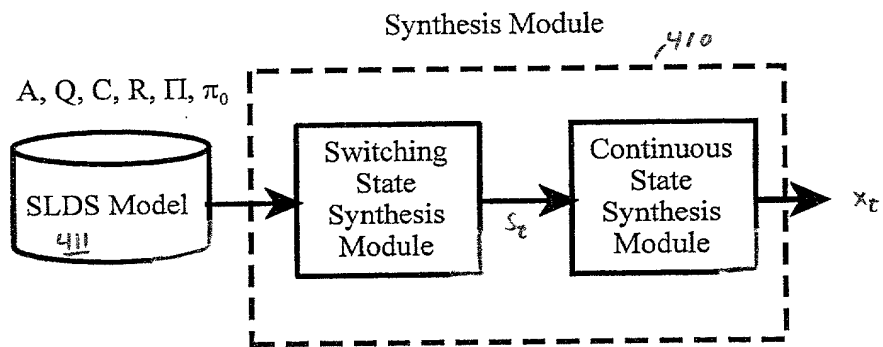


Fig. 20

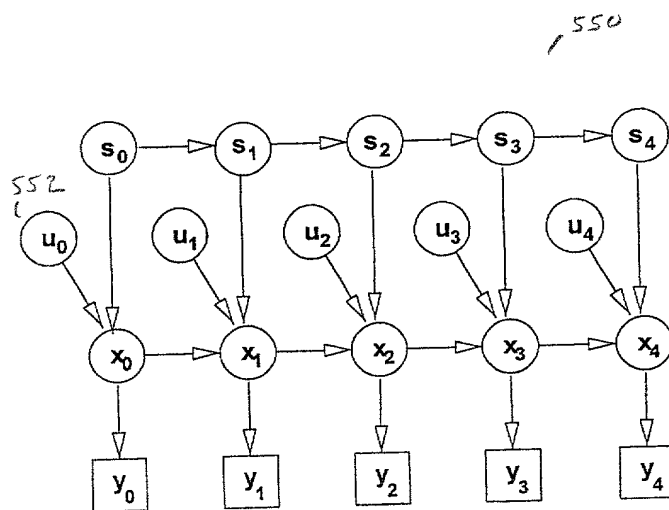


Fig. 21





2.9

580  
/

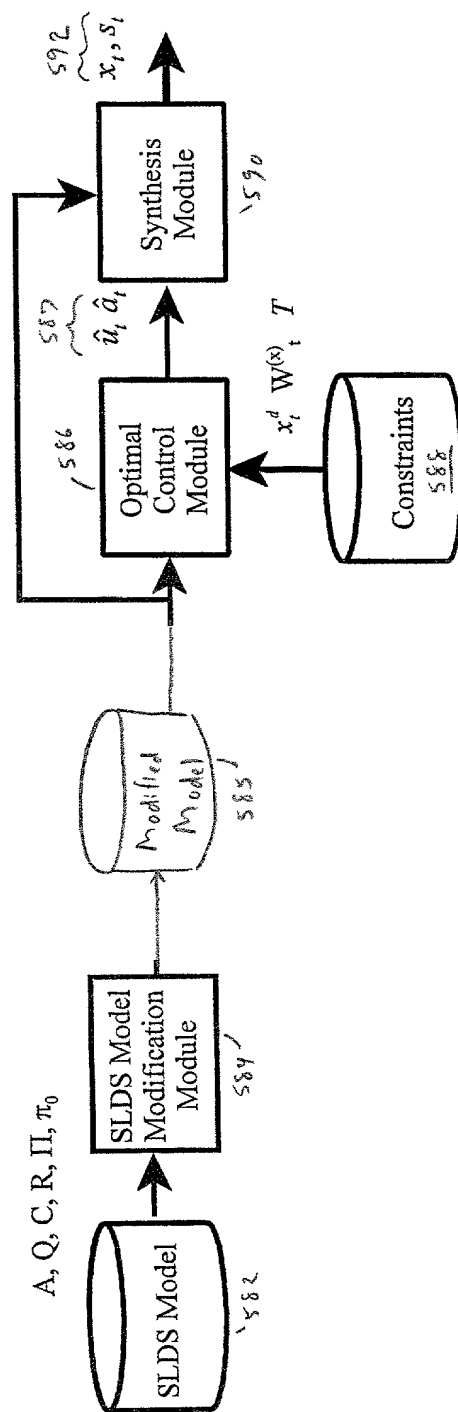


Fig. 24

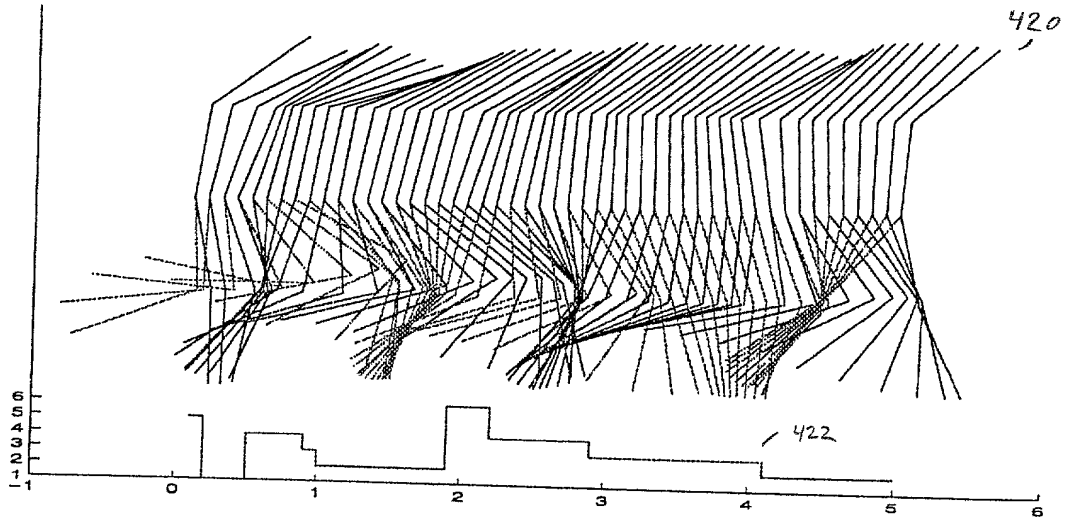
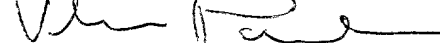



Fig. 25

<b>FULL NAME OF SOLE OR FIRST INVENTOR</b>		<b>INVENTOR'S SIGNATURE</b>	<b>DATE</b>
Vladimir Pavlović			8/29/00
<b>RESIDENCE</b>			<b>CITIZENSHIP</b>
59 Marvin Road, Melrose, Massachusetts 02176			Yugoslavia
<b>POST OFFICE ADDRESS</b>			
(same as above)			
<b>FULL NAME OF SECOND JOINT INVENTOR</b>		<b>INVENTOR'S SIGNATURE</b>	<b>DATE</b>
James Matthew Rehg			8/29/00
<b>RESIDENCE</b>			<b>CITIZENSHIP</b>
106 Quincy Street, Arlington, Massachusetts 02174			USA
<b>POST OFFICE ADDRESS</b>			
(same as above)			

UDMA\MHODMA\Memo;157267;1 Compaq Confidential

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

<b><i>Applicant/Patentee:</i></b>	<b>Vladimir Pavlovic</b>	<b>5</b>
	<b>James M. Rehg</b>	<b>6</b>

**Filed:** \_\_\_\_\_ **\$** \_\_\_\_\_

**Attorney File No.: 0918.1305-000**

**Compaq Ducker Nu.: PD99-2754**

Serial No.:

<b>For:</b>	<b>Method for Motion Synthesis and</b>	<b>\$</b>
	<b>Interpolation Using Switching Linear</b>	<b>\$</b>
	<b>Dynamic System Models</b>	<b>\$</b>

## POWER OF ATTORNEY BY ASSIGNEE

Under the provisions of 37 C.F.R. § 3.71, the undersigned assignee of record of the entire interest in the above-identified patent/patent application by virtue of an assignment (check as applicable):

[X] Concurrently Herewith  
[ ] Date Recorded \_\_\_\_\_  
[ ] Reel \_\_\_\_\_ Frame \_\_\_\_\_  
[ ] Attached Hereto

elects to conduct the prosecution of the application/maintenance of the patent to the exclusion of the inventor(s). The undersigned hereby declares that he/she has reviewed the above-referenced assignment and hereby declares that, to the best of his/her knowledge, title is in the Assignee, and further declares that all statements made herein of his/her own knowledge are true and that all statements made on information and belief are believed to be true. The assignee hereby revokes any previous powers of attorney and appoints the following to prosecute this application/maintain this patent and transact all business in the Patent and Trademark Office connected therewith:

**(Prosecuting Attorney List)**

Irene Kosturakis	33,724
Richard P. Lange	27,296
Barry Blount	35,069
Sarah T. Harris	35,891
Joseph Arrambide	39,589
Keith Lutsch	31,851
Theodore S. Park	26,971

and the attorneys and/or agents associated with Hamilton, Brook, Smith & Reynolds, P.C., Two Millis Drive, Lexington, Massachusetts 02421-4799. Customer No. 21005.

Please direct all communications to: **Hamilton, Brook, Smith & Reynolds, P.C., Two Militia Drive, Lexington, MA 02421-4799, 781-861-6240** to the attention of: **James M. Smith, Esq.**

**ASSIGNEE**

**COMPACT COMPUTER CORPORATION**

Date: SEP 2000

BY: Richard P. Lange  
NAME: Richard P. Lange  
TITLE: Senior Counsel Intellectual Property

**Authorized to Sign this Document on  
Behalf of Compaq Computer  
Corporation**